DOCTOR OF PHILOSOPHY

Analysis of Human Protein Complexes by Quantitative Mass Spectrometry

Kirkwood, Kathryn J.

Award date:
2014

Awarding institution:
University of Dundee

Analysis of Human Protein Complexes by Quantitative Mass Spectrometry

By

Kathryn J. Kirkwood

PhD

University of Dundee

August 2014

Declaration

I declare that I am the author of the thesis; that all references cited have been consulted by me; that the work of which the thesis is a record has been done by me alone or in collaboration with researches named in the text. This work has not been previously accepted for a higher degree.

Kathryn J. Kirkwood

Acknowledgements

First and foremost, I would like to thank Professor Angus Lamond for giving me the opportunity to do this work. His unfaltering supervision and support during my time in the laboratory is greatly appreciated. I also thank Mark who supervised me on a day to day basis. Without his teaching, advice and seemingly endless patience, this project would not have been possible. Similarly, I thank Yasmeen, my kindred spirit in the unfamiliar world of science, for all her help and advice throughout my project. I cannot thank her enough for all the hours she devoted to helping me, both in and out of her university contract. There were many experiments which did not make it into this thesis, in which I was helped by many other Lamond lab members, and to them too, I am incredibly grateful. I thank you all for making my time in the lab such fun, and for providing such wonderful cakes over the years!

I would also like to thank Professor Sara Marshall for giving me the courage to apply for this fellowship. Her mentorship and support throughout my training has been invaluable and I shall never forget the faith she had in me from the start.

Last but not least, I thank my husband, Tom. He too has shown great faith in me and I cannot put into words how grateful I am for everything he has done to support me in this PhD, particularly during the tough times when experiments did not work. Thank you for sticking by me, both in this, and in everything we do. To the remainder of my family and friends, you have all helped my through in your own ways and I could not have done it without you.

Abstract

Proteins are the fundamental building blocks of all living cells and they carry out nearly all cell functions. Proteins predominantly act as part of multiprotein complexes to carry out these functions. The association of protein isoforms and post-translationally modified forms in protein complexes can influence their subcellular location, activity and substrate specificity. It is therefore crucial to characterise protein complexes at the level of the protein isoforms and post-translationally modified forms they contain to fully decipher the network of signalling and regulatory pathways within cells.

The aim of my work has been to develop a technique to study protein complexes through the use of size exclusion chromatography (SEC) combined with tandem mass spectrometry. Combining these approaches has enabled an in-depth analysis of protein complexes in U2OS cells, including those involving post-translationally modified proteins and protein isoforms. The data presented in this thesis provide a proof of concept, together with forming a useful resource, which can be used alongside pull-down analyses to differentiate the interaction partners involved in different protein complexes. This combined approach minimises the need for multiple IP analyses and facilitates a more targeted approach in dissecting the components of individual protein complexes.

I developed this technique further by utilising *in vivo* crosslinking prior to denaturing SEC. This approach enabled more efficient recovery and detection of proteins previously underrepresented using native SEC analysis, including many membrane complexes, thereby providing a more complete picture of endogenous protein complexes.

Abstract

I have applied the native SEC/MS approach in a study of the interactions between the MRFAP1-MORF4L1 proteins. I demonstrated that this complex is distinct from the larger complex involving interactions between the MRGBP-MORF4L1 proteins. In addition, I also demonstrate that the native SEC/MS technique can be extended to assess the effect of drugs on protein-protein interactions.

Overall, the methods I present in this thesis enable the rapid, proteome-wide analysis of endogenous protein complexes, which will advance the future study of protein complexes in biology.

Table of contents

Table of figures

Table of tables

Abbreviations

ADP	adenosine diphosphate
AML	acute myeloid leukaemia
AP-MS	affinity purification-mass spectrometry
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BN-PAGE	blue native-polyacrylamide gel electrophoresis
BS3	bis[sulfosuccinimidyl] suberate
BSA	bovine serum albumin
CBP	calmodulin binding protein
CBQCA	3-(4-Carboxybenzoyl)quinoline-2-carboxaldehyde
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
CID	collision-induced dissociation
CNBr	cyanogen bromide
co-IP	co-immunoprecipitation
DMEM	Dulbecco's modified eagle medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DSP	dithiobis[succinimidyl propionate]
DSS	disuccinimidyl suberate
DTBP	Dimethyl 3,3´-dithiobispropionimidate
DUB	deubiquitinating enzyme
EDTA	ethylenediaminetetraacetic acid
EGFR	epidermal growth factor receptor

EPD	encyclopaedia of proteome dynamics
FCS	foetal calf serum
FDR	false discovery rate
GFP	green fluorescent protein
GO	Gene Ontology
GST	glutathione S-transferase
HCD	higher energy collisional dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HILIC	hydrophobic interaction liquid chromatography
HPA	Human Protein Atlas
HPLC	high performance liquid chromatography
IEX	ion exchange chromatography
IgG	immunoglobulin G
IgM	immunoglobulin M
IP	immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth
LC	liquid chromatography
LC-MS/MS	liquid chromatography - tandem mass spectrometry
LDS	lithium dodecyl sulphate
LysC	endoproteinase LysC
m/z	mass to charge ratio
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NEM	N-ethylmaleimide
NPC	nuclear pore complex

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PPI	protein-protein interaction
PPM	parts per million
pre-mRNA	pre-messenger ribonucleic acid
PTM	post translational modification
RNA	ribonucleic acid
RNase	ribonuclease
RPLC	reverse-phase liquid chromatography
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SILAC	stable isotope labelling by amino acids in cell culture
siRNA	small interfering ribonucleic acid
SMCC	succinimidyl-4-[N-maleimidomethyl] cyclohexane-1-carboxylate
SMPB	N-succinimidyl 4-[4-maleimidophenyl]butyrate
TAP	tandem affinity purification
TBS	tris buffered saline
TBST	tris buffered saline and tween 20
TCEP	tris(2-carboxyethyl)phosphine
TEAB	tetraethylammonium bromide
TFA	trifluoroacetic acid
tRNA	transfer ribonucleic acid
Y2H	yeast two hybrid

Chapter 1: Introduction

All living organisms are made up of cells. From human cells to a unicellular bacterium, proteins form the major component of the machines that perform basic cell functions. Proteins are the building blocks of all living things and they define the structure, function and regulation of cells. Protein is made continuously throughout a cell's lifetime, and is encoded by the heritable message held within DNA as genes. As a result, to understand how cells function at a molecular level, we need to understand what proteins are and what proteins do.

The *proteome* is the entire set of proteins expressed within a cell or organ system and proteomics is the large-scale study of these proteins. As proteins are the key players in cell function, studying the proteins actively expressed in a cell provides the greatest insight into the factors regulating cell biology. Modifications to proteins can alter protein function, localisation, interaction sites and interaction partners - information which would not be apparent by studying genes (the genome) or RNA transcribed from genes (the transcriptome). As will be discussed in more detail in the following section, one gene can code for multiple different proteins therefore the total number of proteins in humans is in vast excess of the ~21, 000 known protein coding genes. It is not only the sheer number of proteins that make the proteome more complex than the genome, but the fact that the proteome differs from cell to cell and varies with time and in response to environmental stimuli increases the challenges faced by the proteomics community.

Proteomics has become synonymous with mass spectrometry and technological advances in this technique over the last decade have opened the door to the in-depth,

high-throughput analysis of proteins. The advances are such that recent studies have been able to demonstrate that there are over 10, 000 proteins in one cell type alone, all identified from one experiment [Beck et al., 2011; Nagaraj et al., 2011]. In the Lamond laboratory, application of mass spectrometry in the study of proteomics has enabled the calculation of global protein turnover rates [Boisvert et al., 2012], comparison of proteins expressed between subcellular compartments, including protein isoforms [Ahmad et al., 2012], analysis of differential protein degradation rates between cellular compartments [Larance et al., 2013] and in the analysis of protein expression during the different phases of the cell cycle [Ly et al., 2014]. By being able to measure such protein parameters on a large scale, we have gained great insight into the regulation of cell function at the molecular level, in both time and space.

1.1 Protein Biology

1.1.1 Protein synthesis

DNA and genes

All living cells store hereditary information in DNA. To optimise the safe and compact storage of DNA, it forms a double helix, which is further compacted by winding around an octamer of histone proteins to form a nucleosome, the basic unit of chromatin (Figure 1.1). Information within DNA relates to protein sequences, RNA sequences and binding domains for RNA molecules and proteins. Therefore, the genome (all the genetic information carried in DNA) contains information not only regarding which proteins to synthesise but also when and where this is to occur.

One of the major milestones of modern science has been the completion of the Human Genome Project, which was developed to map the entirety of the human

genome [Lander et al., 2001]. The human genome is now known to contain ~21, 000 protein coding genes [Clamp et al., 2007] and although greater knowledge of the genome provides an understanding of the genetic defects in some diseases [Lander, 2011] it does not always provide the greatest understanding behind cell biology as genes themselves are non-functioning. Although the genetic code is generally identical from one cell to the next in humans (excluding special cases, such as plasma cells, gametes and tumour cells), it is known that the *expression* of genes varies greatly from one cell type to another and even within one cell type, gene expression can vary at different points in time. Modification of histone molecules within chromatin, for example by histone acetylation and deacetylation complexes, can act as a signal for either chromatin relaxation, or chromatin compaction, respectively, thereby defining whether protein coding genes held within DNA can be transcribed at a specific point in time. Unravelling of the chromatin and opening of the DNA helix permits DNA-binding proteins (transcription factors) and RNA polymerase II to bind specific DNA sequences (the promoter) to initiate gene transcription (Figure 1.1).

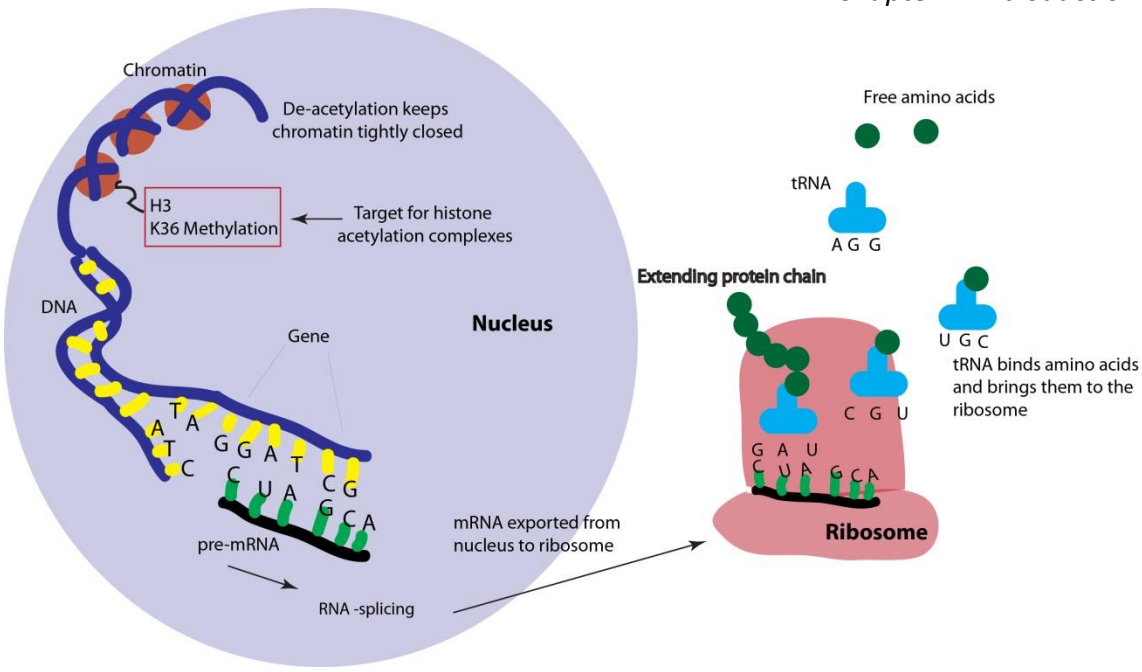


Figure 1.1 Schematic overview of protein synthesis

RNA to protein

Transcription and translation are the processes by which an RNA product or protein is made. As the RNA polymerase II passes along the DNA, it forms pre-messenger RNA (pre-mRNA). There are more mRNA transcripts than there are genes in the human genome and this is, in part, due to RNA splicing. Pre-mRNA contains coding (exons) and non-coding (introns) sequences and in order for the sequence to be accurately translated into a protein, the non-coding sequences need to be removed to form mRNA. Variation in this process can lead either to the formation of different versions of the same protein (protein splice isoform), or to an entirely new protein altogether, all encoded from the same gene. Splicing may create or remove a protein interaction interface for binding interaction partners, thereby increasing protein functional diversity. The resultant mRNA is then transferred from the nucleus to the cytoplasm where protein synthesis takes place in ribosomes. Here, the mRNA code is read by transfer RNA (tRNA), and the encoded amino acids are linked to form a protein sequence.

Transcriptomics is the study of the set of RNA molecules transcribed from genes, either in a cell at a given point in time, or in response to different environmental stimuli. However, evidence suggests that not all mRNA is translated into protein and several studies report that mRNA levels do not uniformly reflect protein levels [Gygi et al., 1999b; Maier et al., 2009; Lundberg et al., 2010; Vogel et al., 2010; Nagaraj et al., 2011; Schwanhausser et al., 2011; Ly et al., 2014]. Although two studies do show a high correlation between mRNA and protein levels, statistical analysis does not suggest this correlation is high enough to reliably use mRNA levels as a predictor for protein expression [Lundberg et al., 2010; Ly et al., 2014]. Such discrepancies could be due to differences in translational efficiency [Man and Pilpel, 2007; Spruill and McDermott, 2009], post-translational modification of proteins [Ly et al., 2014], or to variations in the rate of protein degradation/half-life [Varshavsky, 1996]. This evidence highlights the importance of studying cell biology at the protein level for a greater understanding of cell function.

1.1.2 Protein Regulation

Overall protein quantity is regulated at the level of both production and degradation. It is known that proteins vary widely in their turnover rates, with some being present for only a few minutes (e.g. p53) [Haupt et al., 1997] and others surviving for more than a month or even longer (e.g. histones and nuclear pore complex components) [Toyama et al.], although the majority of proteins, in the setting of cell culture, have a turnover in the region of 20 hours [Boisvert et al., 2012]. Not only is the rate of protein synthesis important in the regulation of protein concentrations but the rate of protein degradation also plays a significant role. Various processes are involved in both, and these are summarised in Figure 1.2.

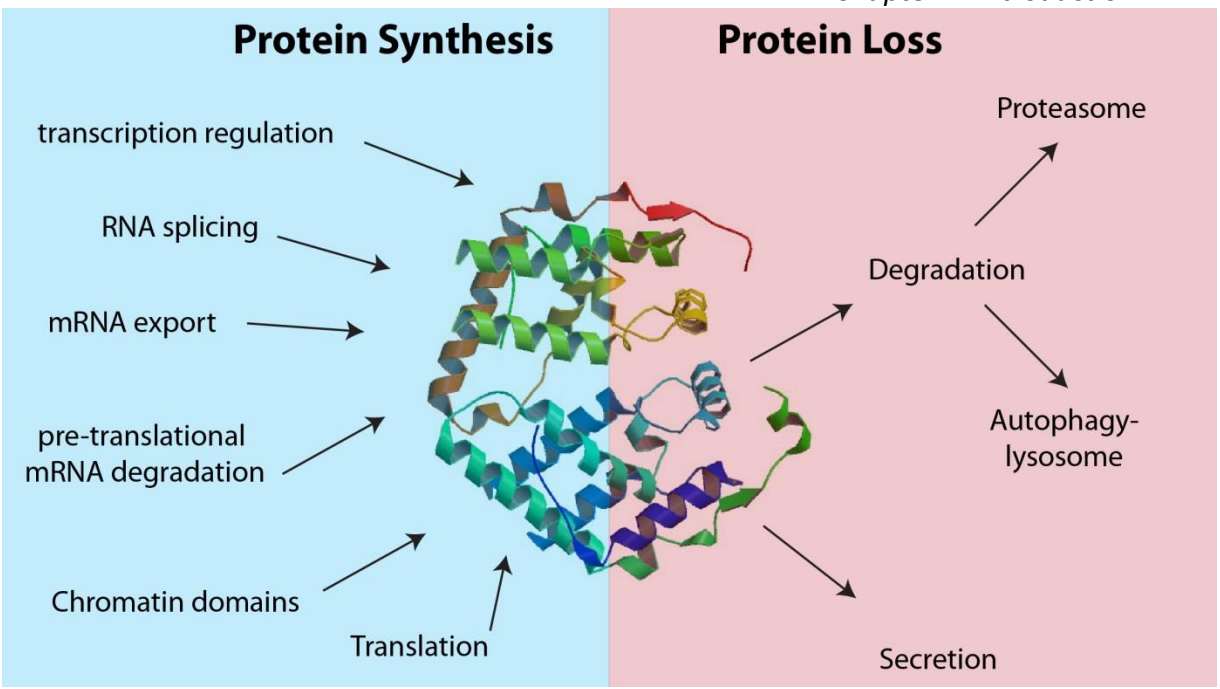


Figure 1.2. Factors influencing protein regulation.

Regulation of protein synthesis

As all cells in an organism contain the same DNA, it is this regulation of gene expression that gives a cell control over its structure and function, and forms the basis of cellular differentiation. The processes of transcription and translation have been described previously in this section but there are many factors which control the rate at which genes are expressed and proteins synthesised to enable a cell to respond accurately to changes in cell state and/or environment. These rate controlling steps occur throughout the pathway from DNA to protein and include: regulating when and how often a mRNA is transcribed; altering the splicing and processing of mRNA; regulating which mRNAs are translated by the ribosome and their rates of translation; degrading some mRNAs in the cytoplasm prior to translation, regulating rates of protein degradation; and by modifying the final protein product to alter functional

capacity e.g. by either cleavage or by post translational modification. In some cases, regulation of gene transcription is particularly important during certain situations.

Transcription rates can be affected by proteins binding directly to DNA at specific sites around the coding region (i.e. promoter and enhancer sites). Such events may block RNA polymerase binding sites, or mask activation surfaces of activating proteins, thereby preventing a gene being transcribed. For example, conjugation of SUMO2 to specificity protein 1 (Sp1), a major transcription factor, attenuates DNA binding, impairs the interaction between Sp1 and the co-activator p300, and leads to the recruitment of the repressor, Sp3 to gene promoters, thereby negatively regulating gene expression [Gong et al., 2014]. Alternatively, the binding of some proteins to DNA can enhance transcription by assisting RNA polymerase binding. For example, transcription initiation by RNA polymerase II requires the co-ordinated interaction of several transcription factors to form the preinitiation complex at the core promoter [Nikolov and Burley, 1997]. This complex facilitates the recruitment of RNA polymerase II to DNA and therefore enhances transcription.

Transcription factors themselves can be phosphorylated, in addition to other modifications (glycosylation, acetylation, ubiquitination and sumoylation), inhibiting their ability to either bind DNA, recruit RNA polymerase II, or promote transcription elongation [Tootle and Rebay, 2005]. In some cases, activation of a signalling cascade reverses the modification process, thereby activating the transcription factor and facilitating transcription, in response to a cellular stimulus. One example of this process has been described by Okamura *et al.*, where dephosphorylation of the NFAT1 transcription factor masks a nuclear export signal, exposes a nuclear localisation signal, and promotes transcriptional activity [Okamura et al.].

Modifications to histone tails, such as methylation, can facilitate additional histone modifications (acetylation or deacetylation) which either promote, or suppress, gene expression through the regulation of chromatin structure. Methylation, acetylation and deacetylation processes can involve the function of a common protein. One example of this is MORF4L1. This protein has a N-terminal chromodomain through which it binds di/tri methylated histone H3 on K36, and an MRG domain at its C-terminus (Figure 1.3) [Zhang et al., 2006a]. The MRG domain is known to interact with components of histone acetylation [Cai et al., 2003] and histone deacetylation complexes [Kumar et al., 2011] thereby regulating chromatin structure and gene expression (Figure 1.3A,B). In addition, this protein has been demonstrated to have a role in regulating methylation of histone H3K4 through an interaction with RBP2 [Hayakawa et al., 2007] (Figure 1.3C). Consequently, factors affecting MORF4L1 binding to its interaction partners will affect gene expression and transcription.

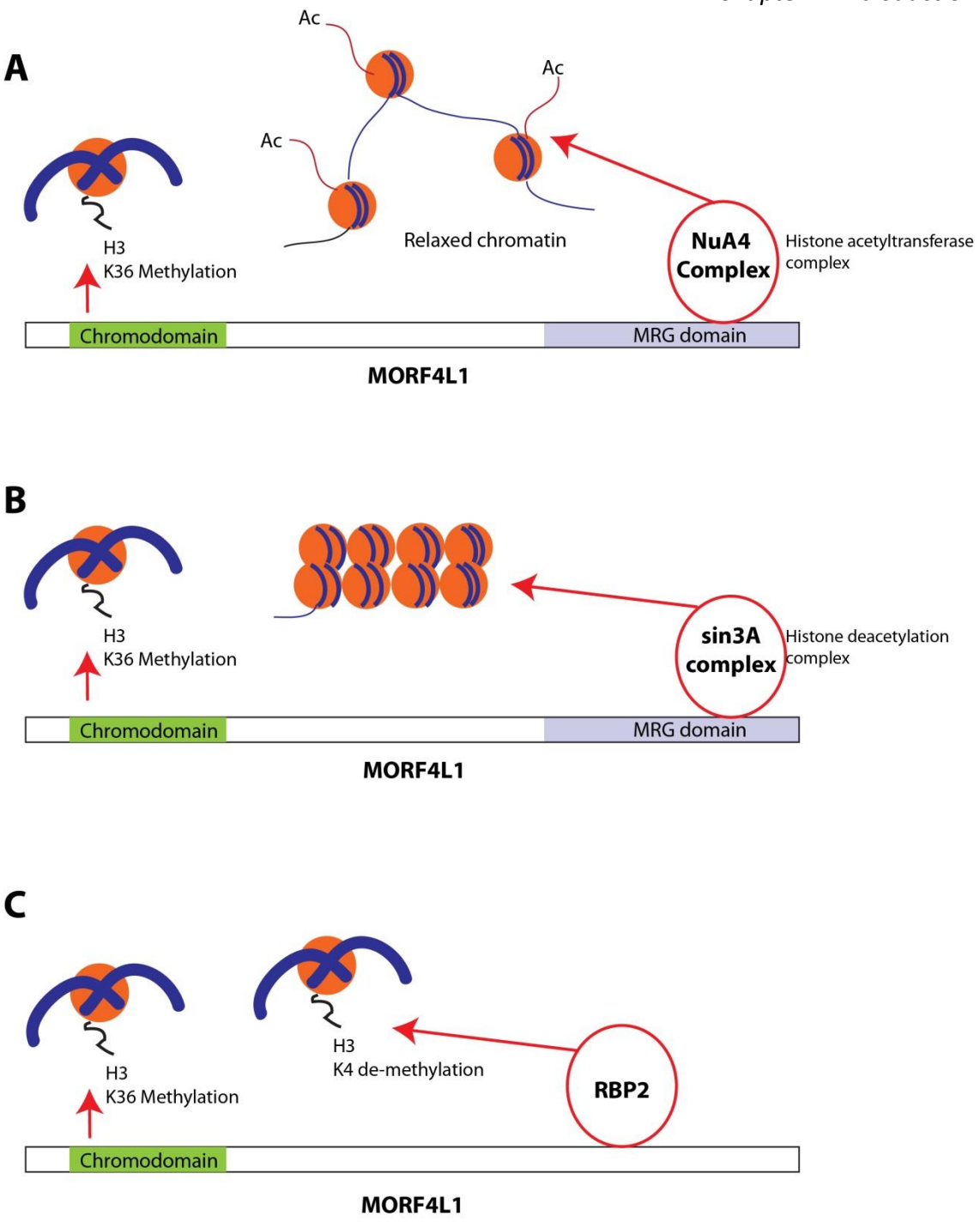


Figure 1.3. Schematic diagram of the MORF4L1 protein and its role in the regulation of protein synthesis.

The MORF4L1 protein binds to methylated tails of Histone H3 via its chromodomain. The MRG domain for MORF4L1 is the binding domain for many interacting proteins. *A*, the NuA4 histone acetyltransferase complex binds the MRG domain to facilitate the acetylation of histone H4 which opens out chromatin. *B*, Histone deacetlyation complexes (e.g. sin3A complex) also bind the MRG domain leading to chromatin compaction. *C*, RBP2 interacts with MORF4L1 to regulate methylation of Histone H3.

It is clear to see how the alteration of any factor which changes the expression levels, structure or function of any protein involved in transcription activation could have huge implications on the rate of transcription and therefore protein synthesis.

Protein loss and degradation

Aside from transcription, loss of protein expression obviously also affects total protein levels. Some proteins are secreted and therefore intracellular levels are reduced however degradation of proteins is a major factor in protein regulation. There are two main pathways for intracellular protein degradation i.e. the proteasome and autophagy-lysosomal systems (Figure 1.2). In general, short lived proteins such as those involved in cell cycle regulation, are degraded by the proteasome and longer living proteins are degraded by the lysosome [Ohsumi, 2006].

Many copies of the proteasome are present throughout the cytosol and nucleus of the cell. Proteins that have not formed properly or those that fail to fold are rapidly degraded following synthesis. Other proteins which must be quickly regulated in response to different cell sates can also be degraded via this pathway. The 26S proteasome is an ATP-dependent protease comprising a core (20S) particle and a 19S regulatory cap [Peters et al., 1993; Adams, 2003]. The regulatory cap is made up of at least 19 subunits [Sharon et al., 2006], several of which are ATPases, believed to be important for unfolding proteins before they are passed into the cylindrical core where proteolysis takes place. The regulatory cap is involved in recognising proteins which contain a tag or modification indicating that they have been targeted for degradation (including the K48-ubiquitin chain) [Clague and Urbé, 2010]. These markers for degradation can be added to a protein as the end result of a signalling pathway [Glickman and Ciechanover, 2002]. Some proteins are continuously degraded after

synthesis as part of a biological process, such as the tumour suppressor gene p53. This rapid degradation results in only a low level of the protein under normal growth conditions. However, oncogene activation leads to sequestration of the E3 ligase Mdm2 and thus prevents p53 degradation and as a result, there is a rapid increase in p53 levels (Figure 1.4) [Haupt et al., 1997]. This demonstrates how regulation of protein degradation can be used to rapidly respond to altered cell states independently of transcription activation. MG132 is an inhibitor of the 26S proteasome and it has been used for the study of this degradation pathway [Tatham et al., 2011; Leidecker et al., 2012; Larance et al., 2013].

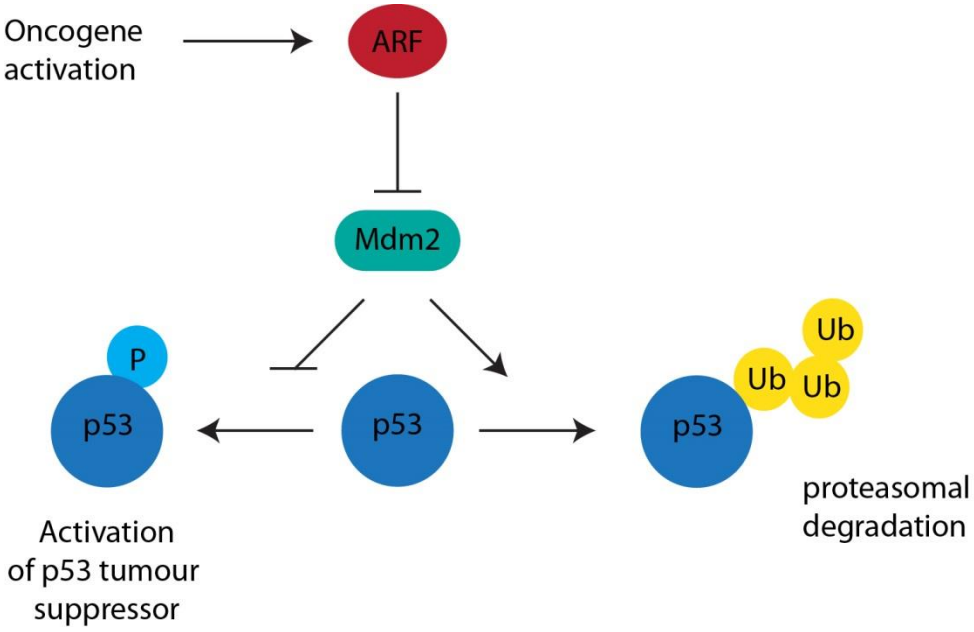


Figure 1.4. Schematic diagram of p53 regulation by the ARF/Mdm2 pathway.

Lysosomes are membrane bound organelles which are involved in the destruction of proteins, nucleic acids, lipids and other cellular debris. Material is delivered to the lysosome in vesicles where lysosomal hydrolases break down digested molecules. Endocytolysis involves the process of taking up macromolecules from extracellular

fluid, in vesicles, for degradation. A second pathway involving the lysosome however is autophagy [Feng et al., 2014]. This pathway for degradation involves cytoplasmic proteins and organelles of the cell itself becoming enclosed in a double membrane, thus creating an autophagosome, which then fuses with a lysosome for degradation, forming the autolysosome. Similarly to the proteasome, amino acids resulting from lysosomal degradation can be re-used within the cell for future protein synthesis [Feng et al., 2014]. The autophagosome is not always present in a cell – it forms quickly and disappears under the regulation of autophagy activating signals [Lamb et al., 2013]. Dysregulation of the autophagy pathway has been implicated in some human diseases, such as Crohn's disease [Murthy et al., 2014], cancer and neurodegeneration [Choi et al., 2013] and a knockout mouse model has demonstrated that autophagy is essential for post-natal survival [Kuma et al., 2004]. Although autophagy has been described as a non-selective degradation process, ubiquitin modification has also been shown to direct degradation via this pathway [Clague and Urbé, 2010]. Modification of proteins by a ubiquitin side chain (K63-linked ubiquitin chain) is recognised by p62 and NBR1 and therefore targeted for autophagy-mediated degradation [Kirkin et al., 2009]. Evidence suggests there is crosstalk between the proteasomal and autophagy regulation pathways. Which pathways are selected for different proteins may relate to subcellular localisation, ubiquitin chain length and linkage type [Clague and Urbé, 2010; Driscoll and Chowdhury, 2012].

1.1.3 Protein structure

Amino acids are a family of 20 molecules within a cell that have a similar basic structure. These molecules all have an amino group (H_2N) at one end termed the N-terminus, and a carboxylic acid group (COOH) at the other, termed the C terminus.

Linked to both these termini is a carbon atom, termed the α-carbon, which attaches to a side chain (Figure 1.5).

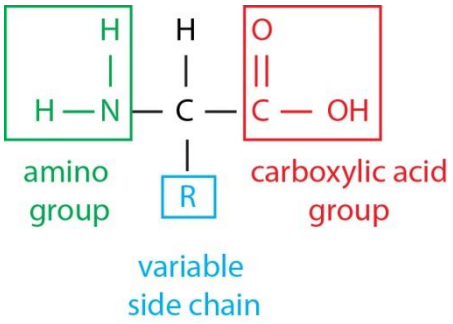


Figure 1.5. Basic structure of an amino acid

Each side chain can have different properties; five form ions in solution at physiological pH and therefore carry either a positive (histidine, lysine and arginine), or negative charge (aspartic acid and glutamic acid). Some of the side chains are hydrophilic and others are hydrophobic (Table 1.1). It is this variability in the side chain that distinguishes the chemical properties of the 20 different amino acids. Amino acids can link together, the C-terminus of one amino acid binding to the N-terminus of the next, forming a peptide bond between the two. The linear sequence of amino acids is known as the protein **primary structure** (Figure 1.6).

Amino Acid	Abbreviation (3 letter)	Abbreviation (1 letter)	Side chain property
Lysine	Lys	K	Basic
Arginine	Arg	R	Basic
Histidine	His	H	Basic
Aspartic acid	Asp	D	Acidic
Glutamic acid	Glu	E	Acidic
Asparagine	Asn	N	Uncharged polar
Glutamine	Gln	Q	Uncharged polar
Alanine	Ala	A	Nonpolar
Valine	Val	V	Nonpolar
Leucine	Leu	L	Nonpolar
Isoleucine	Ile	I	Nonpolar
Proline	Pro	P	Nonpolar
Phenylalanine	Phe	F	Nonpolar
Methionine	Met	M	Nonpolar
Tryptophan	Trp	W	Nonpolar
Glycine	Gly	G	Nonpolar
Cysteine	Cys	C	Nonpolar

Table 1.1 Properties of amino acids

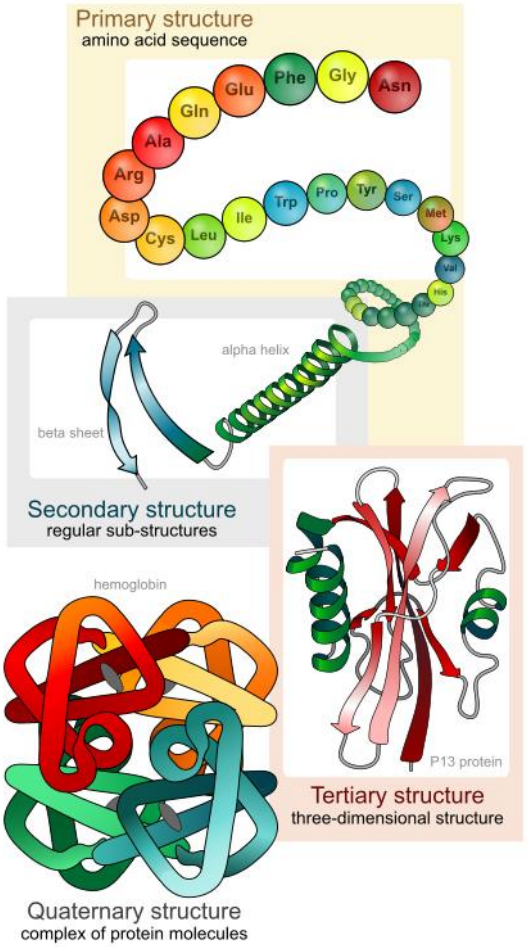


Figure 1.6. Primary, secondary, tertiary and quaternary protein structures.

(http://en.wikipedia.org/wiki/Protein_structure#mediaviewer/File:Main_protein_structure_levels_en.svg)

In addition to the peptide bonds between amino acids, hydrogen bonds can also form between the main-chain peptide groups. These can cause a protein to fold in certain ways to form highly regular local sub-structures such as the alpha-helix and the beta sheet. These substructures are referred to as the protein **secondary structure** (Figure 1.6).

Chains of typically between 50 to 2000 amino acids then fold up in an organised fashion, dependent on the properties of the amino acid side chains, to form the 3D structures of proteins, known as the **tertiary structure** (Figure 1.6). Proteins will fold in the most energy efficient way to accommodate opposing forces of hydrophobic and hydrophilic amino acids. The addition of some solvents can lead to unfolding (denaturing) of proteins by disrupting the non-covalent interactions, but many proteins will spontaneously re-fold again after this has been removed. In other cases protein folding requires the action of protein chaperones to ensure it occurs correctly. The **quaternary structure** is the 3D structure of a multi-subunit protein, held together by the same forces which regulate the tertiary structure.

Although each protein contains a unique sequence of amino acids, some large proteins contain amino acid sequences which are capable of folding largely independently of the remainder of the protein. These sequences can form structural domains and although they can be made of different linear combinations of amino acids, if the overall properties are similar, it can lead these domains to act in predictable way. These are often the modular units which facilitate the different functions of a protein (for example, the chromodomain and MRG domain of MORF4L1 (Figure 1.3). Mapping the structural domains can help decipher the probable structure of a protein [Kelley

and Sternberg, 2009], and protein domains have been used to determine possible protein-protein interactions [Wojcik and Schächter, 2001; Memisevic et al., 2013].

1.1.4 Protein modifications

As discussed previously, the complexity of the human proteome is much greater than that of the genome. Even more complexity is added to the proteome at a post-translational level through proteolytic cleavage and by the addition of post-translational chemical modifications (PTMs) of the amino acid side chains. Such modifications alter the structural and chemical properties of the amino acids and thereby can alter protein function (Figure 1.7).

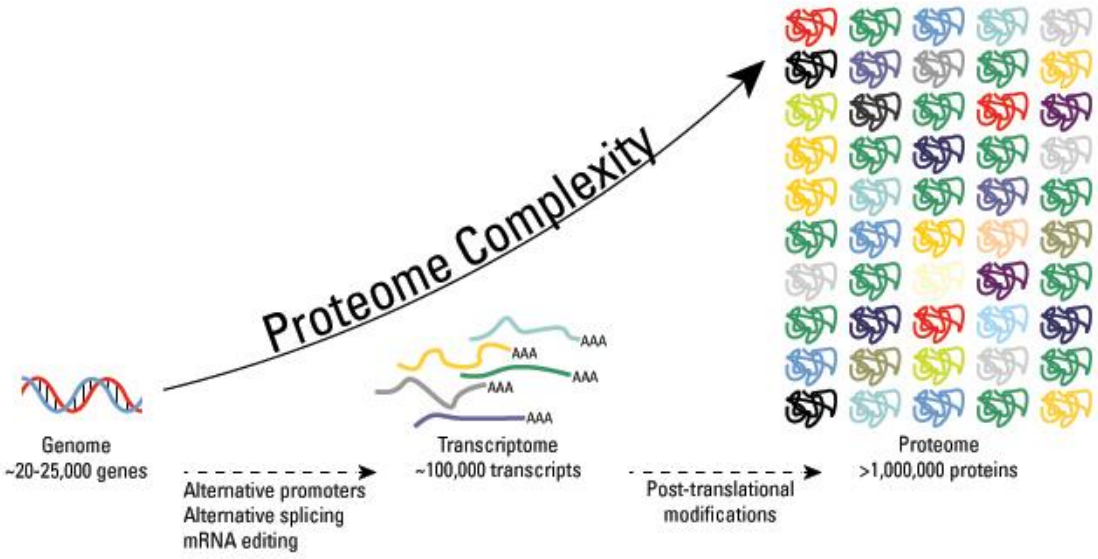


Figure 1.7. Post-translational modifications increase the complexity of the proteome.

http://www.piercenet.com/method/overview-post-translational-modification

A post-translational modification of a protein refers to the conjugation of a small chemical group (e.g. PO_4, OH, CH_3 etc.) or even a small peptide (e.g. SUMO, NEDD8, ubiquitin) to the side chain of an amino acid in a protein to increase functional

diversity. PTMs offer a very rapid and dynamic way to regulate protein activity and they do this in many ways, for example by altering protein-protein interactions by altering binding sites, inducing a conformational change to alter protein function and altering protein localisation (Figure 1.8). PTMs are catalysed by enzymatic activity and many are reversible with separate enzymes used to add or remove the modification (e.g. kinases and phosphatases). There are a vast number of different types of PTM, many of which may be detected at multiple residues. Furthermore, multiple PTMs can be present on the same protein, further increasing combinatorial complexity [Khoury et al., 2011]. Some of the common PTMs are described in the following paragraphs.

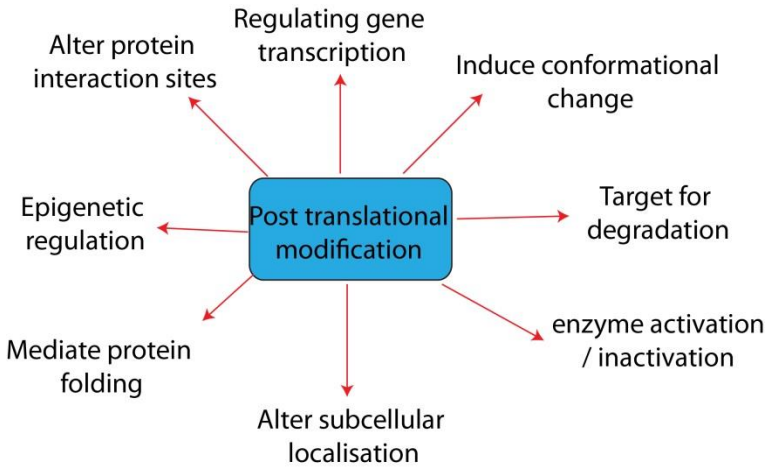


Figure 1.8. Summary of the biological effects of post-translational modifications

Phosphorylation

One of the most widely studied PTMs is phosphorylation, which involves the addition of a phosphate group by a kinase to the hydroxyl group on the side chains of serine, threonine and/or tyrosine residues. This modification is reversible, and is removed by phosphatases (Figure 1.9). The addition or removal of a phosphate group can change the strength of physical protein-protein interactions, affect protein stability and alter

protein dynamics (reviewed by [Johnson, 2009; Nishi et al., 2011]). Phosphoproteomic analyses have demonstrated that most mammalian proteins can be phosphorylated and therefore phosphorylation has a vast role in protein regulation [Olsen et al., 2006; Olsen et al., 2010]. Phosphorylation plays critical roles in the regulation of many cellular processes including regulation of metabolism, cell signalling and cell migration. For example, syndecan-4 phosphorylation has recently been identified as a key regulator of focal adhesion dynamics and cell migration through the regulation of integrin recycling [Morgan et al., 2013]. The reversibility of phosphorylation makes this PTM eminently suitable for signal transduction, which allows a cell to react rapidly in response to intracellular or extracellular stimuli. A protein sensing a stimulus as a result of ligand binding, protein cleavage or any other stimulus can then activate second messengers and signalling enzymes through the activation of protein kinases. The signal amplification resulting from such an enzyme cascade can lead to the activation of a global cellular function from one molecule. One example of this is the activation of cell proliferation, via the ERK module, by a single growth factor [Cowan and Storey, 2003]. Signalling pathways involving TOR (a protein kinase) are also known to regulate growth related processes [Wullschleger et al., 2006]. In addition, TOR also controls many aspects of cell metabolism, including amino acid biosynthesis and glucose homeostasis, and inhibitors of this pathway have been highlighted as being of potential benefit for the treatment of cancer, cardiovascular disease, autoimmunity and metabolic disorders [Wullschleger et al., 2006].

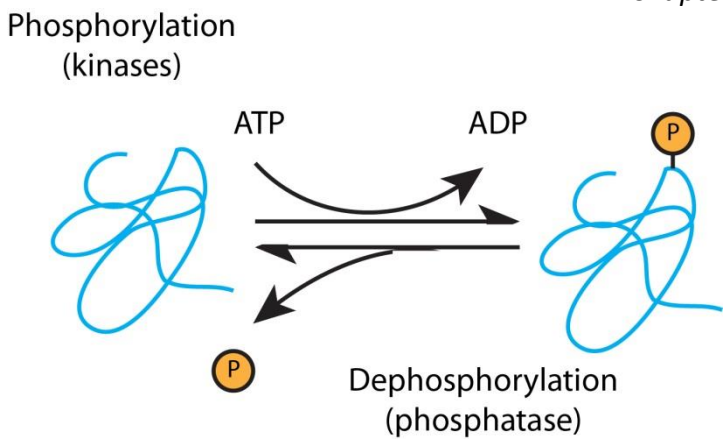


Figure 1.9. Schematic diagram of phosphorylation.

Phosphate groups are added and removed from a protein by kinases and phosphatases, respectively.

Ubiquitination

Ubiquitin is a small peptide which can be conjugated to lysine residues in proteins [Pickart and Eddins, 2004]. The ubiquitination process occurs in three stages: activation, conjugation and ligation and these stages involve ubiquitin activating enzymes (E1), ubiquitin conjugating enzymes (E2) and ubiquitin ligases (E3) respectively (Figure 1.10) [Pickart and Eddins, 2004]. The result is the conjugation of ubiquitin to lysine residues or to the amino group of a protein's N-terminus [Pickart and Eddins, 2004]. Ubiquitination, like many PTMs, is reversible, a process facilitated by deubiquitinating enzymes (DUBs) [Wilkinson, 1997]. As described in the previous section, ubiquitination has a major role regulating protein degradation via the proteasome and autophagosome [Clague and Urbé, 2010]. The formation of polyubiquitin chains is a key factor in this regulatory process, with ubiquitin chains at lysine 48 being recognised as a degradation signal by the proteasome [Clague and Urbé, 2010], and ubiquitin chains at lysine 63 being targeted for autophagy mediated degradation [Kirkin et al., 2009]. Other types of ubiquitin conjugates are involved in

the regulation of many different cellular processes, independently of proteolytic degradation, including gene transcription, DNA repair and replication, and intracellular trafficking [Haglund and Dikic, 2005]. This regulation may occur through the ability of ubiquitin to mediate protein-protein interactions, its role in receptor endocytosis and its role in the regulation of NF-κB signalling [Haglund and Dikic, 2005]. Ubiquitination can induce conformational changes in proteins to alter subcellular localisation. For example, ubiquitination of p53 induces a conformational change which exposes a nuclear export signal. p53 is subsequently expressed in the cytosol where it possibly inhibits transcriptional activity and may promote cytoplasmic apoptotic function [Carter et al., 2007].

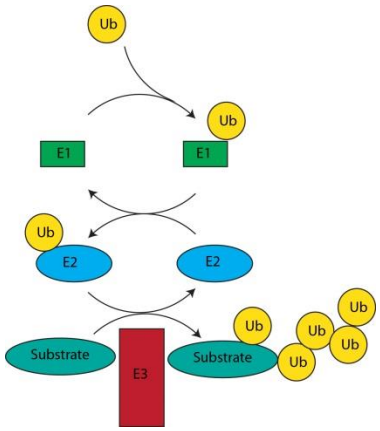


Figure 1.10. Schematic diagram of ubiquitination.

Ubiquitin is activated in a two-step reaction involving the E1 ubiquitin-activating enzyme. E2 ubiquitin conjugating enzymes catalyst the transfer of ubiquitin from the E1 enzyme to the active site on the E2 enzyme. E3 ubiquitin ligases catalyse the final step in the ubiquitination cascade in which the ubiquitin is transferred to the target protein.

Methylation

Protein methylation involves the addition of a methyl group, typically to arginine or lysine residues of a protein. Arginine can be methylated once or twice whereas lysine

can be methylated up to three times [Kouzarides, 2007; Zhang et al., 2012]. The majority of the information available about methylation concerns its role in histone modification and consequently gene expression. In general, methylation of histone H3 lysine residues K4, K36 and K79 are linked to activation of transcription, whereas di and tri-methylation on H3K9, H3K27 and H4K20 are associated with gene silencing [Kouzarides, 2007; Zhang et al., 2012]. Histone lysine methylation has also been associated with other chromatin-associated processes including replication [Rivera et al., 2014] and DNA-repair [Kouzarides, 2007]. Although histones are the predominant proteins modified by methylation, evidence also suggests that some non-histone proteins including: p53, transcription factors and DNA-modifying enzymes are methylated [Zhang et al., 2012].

Acetylation

Acetylation is a PTM in which an acetyl group is added to proteins, classically to the amino group on the side chain of lysine residues. This reaction relies on acetyl-coenzyme A as the acetyl donor group and is catalysed by acetyltransferases (Figure 1.11). As described in section 1.1.2, acetylation of lysine residues in histone tails is a vital part of gene regulation as this opens chromatin to facilitate transcription [Kuo and Allis, 1998]. Deacetylation is the removal of the acetyl group and in the context of histone deacetylation, proteins involved in this process inhibit gene transcription by helping chromatin to return to its closed state [Kuo and Allis, 1998].

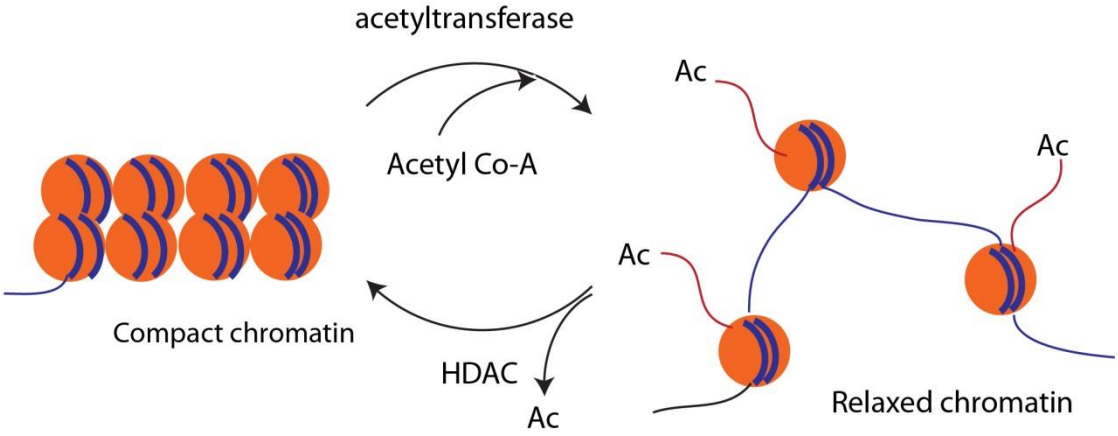


Figure 1.11. Schematic diagram of acetylation.

Acetyltransferases catalyse the addition of an acetyl group onto a target protein and deacetylation complexes perform the reverse action.

NEDDylation

NEDDylation is a very similar process to ubiquitination in that it involves the conjugation of a small ubiquitin like peptide, NEDD8, to substrate proteins through an enzyme cascade involving an E1 activating enzyme, E2 conjugating enzyme and E3 ligase. Known NEDD8 substrates include cullins, p53, p73, Mdm2, pVHL, BCA3, EGFR, APP and L11 [Rabut and Peter, 2008; Xirodimas, 2008; Watson et al., 2011]. NEDDylation is known to induce conformational changes, preclude certain interactions and provide a novel binding surface (reviewed by [Rabut and Peter, 2008]).

Stoichiometry of PTMs

Overall abundance is not always an accurate indicator of the functional activity of a protein, which is why it is important to analyse proteins at the level of PTMs. For example, the overall abundance of a protein may be stable but the constant addition and removal of a PTM in response to changing cellular or environmental stimuli can activate or deactivate the functional capacity of that protein. Therefore, not only are

the different sites of modifications on a peptide of interest, but knowing the stoichiometry of a specific PTM at each site on a protein is important and can provide clues as to its role and functional significance [Olsen and Mann, 2013]. PTM stoichiometry relates to the proportion of modified and unmodified residues present at the site of modification in a protein, and reflects the balance between rates of enzymatic PTM addition and removal, which can vary over time and thereby regulate cell function. For example, the stoichiometry of cyclin-dependent kinase phosphorylation is known to vary throughout the cell cycle, reflecting the regulatory role this protein has during mitosis [Olsen et al., 2010]. High PTM stoichiometry has been reported to be a good indication that a site may be functional [Olsen and Mann, 2013]. However, it is also acknowledged that the converse does not hold true, because even a small pool of protein that bears a modification may specifically correlate with a subset of protein that varies in activity over time and /or space when compared to the non-modified form across the entire cell.

The stoichiometry of a PTM required to complete a function in a cell will vary depending on the type of modification, the site of the modification, the subcellular localisation and the environmental stimulus, in addition to countless other parameters. Ongoing research in this field will provide invaluable information regarding cell regulation and function.

Almost every aspect of cell biology and pathogenesis is affected by PTMs and the diversity of effects conferred on a protein as a result of PTMs demonstrates why knowledge of proteins at the post-translational level is vital for a greater understanding of cell biology. Altered expression or activity of a protein can be the

trigger for a wide range of human diseases therefore it is hardly surprising that pharmaceutical companies are targeting PTM components for drug development strategies [Cohen, 2009; Cohen and Tcherpakov, 2010].

1.2 Protein Analysis

1.2.1 Mass spectrometry as a tool to study proteins

Mass spectrometry is a technique used in the proteomics field to separate and identify the peptides contained in a complex protein mixture on the basis of their mass-to-charge ratio. This is achieved through the use of a mass spectrometer, which utilises electromagnetic fields to control the movement of charged molecules (ions) within the instrument. A total cell lysate contains a huge number of proteins which is a very complex mixture, even more so at the peptide level. To reduce the complexity of a sample prior to injection into a mass spectrometer, samples can be separated into multiple fractions, usually by reverse phase liquid chromatography (RPLC). RPLC utilises peptide affinity for non-polar beads (packed in a column) in a buffer which varies from mostly aqueous (2% acetonitrile in water) to mostly organic, non-polar (80% acetonitrile) over the time it takes for a sample to be run. The beads have linear octadecane groups (C18) attached to the surface by covalent bonds. C18 groups are hydrophobic (non-polar) and can bind polar molecules (i.e. charged peptides) in a highly polar solvent such as water. The polarity of the solvent is then reduced by increasing the concentration of acetonitrile. As the solvent polarity matches or exceeds the non-polarity of a molecule bound to C18, the molecule will elute into the buffer flowing through the column at that time. Different molecules will elute at different acetonitrile concentrations as their varying polarity will require different degrees of non-polarity to elute from the beads. This facilitates sample separation prior to

injection into the mass spectrometer. Mass spectrometric measurements are performed in the gas phase on ionised peptides. To generate either negatively or positively charged gas phase peptide ions from the eluate of RPLC, electrospray ionisation is usually used as the ion source. The charged peptides are then passed to a mass analyser, which separates ionised analytes on the basis of their mass-to-charge ratio (m/z) and then they pass to a detector which registers the number of ions with a specific m/z ratio. Figure 1.12 describes the basic structure of a mass spectrometer.

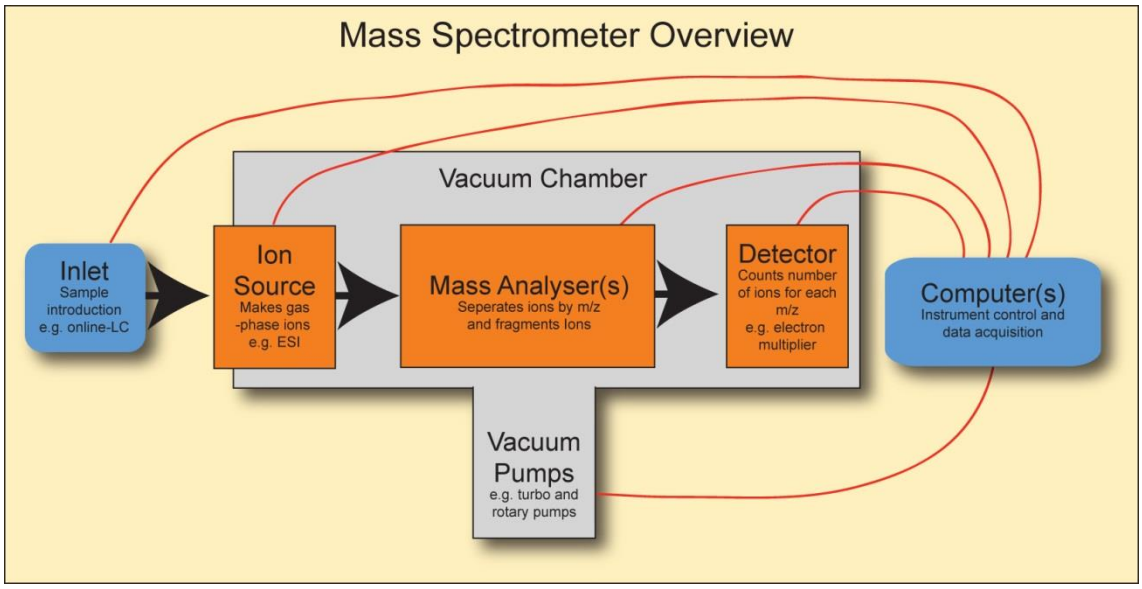


Figure 1.12. Overview of a mass spectrometer.

(http://www.lamondlab.com/MSResource/index.php)

To enable ions to move as freely as possible, the mass spectrometer is maintained in a vacuum chamber. Once passed from the ion source into the chamber, the rate of movement of an ion is controlled by its mass (m) and its charge state (z) and the mass spectrometer measures these in a mass-to-charge ratio (m/z). The separation of ions in the instrument takes place in the mass analyser which separates the ions on the basis of their m/z ratio to permit their separate detection. There are four basic types of mass

analyser used in the proteomics field: ion trap, time of flight (TOF), quadrupole and Fourier transform ion cyclotron [Aebersold and Mann, 2003]. These analysers can either be used individually, or combined, to take advantage of the strength and weaknesses of each approach in their powers of resolution, sensitivity and detecting mass accuracy.

The mass spectrometers used in this thesis were the LTQ Orbitrap Velos and the Q Exactive, both from Thermo Scientific Fisher. The Velos system uses a combination of a low resolution linear ion trap and an orbital ion trap. The linear ion trap performs the initial scan to identify charged peptides before sending them to the orbital ion trap for ion fragmentation. The orbital ion trap radially traps ions between a central spindle electrode and an outer barrel-like electrode. The ions orbit the central spindle and oscillate along the axis of the electric field. The frequency of the ion oscillations allows a mass/charge value to be calculated. By combining the two forms of ion trap, the instrument provides better mass resolution, high mass accuracy, large dynamic range and high mass/charge range [Hu et al., 2005]. The Q Exactive is a hybrid quadrupole-Orbitrap instrument that does not have a linear ion trap. Therefore, the Orbitrap is the main mass analyser for this instrument.

Mass-to-charge ratio data is used in the generation of a mass spectrum, which shows a graph of the relative intensity of all ions detected in the mass spectrometer within a defined period of time as a function of their m/z ratio. To convert the peptide spectral data obtained from the mass spectrometer into protein identifications, specialist software is used, such as Mascot and Andromeda [Perkins et al., 1999; Cox et al., 2011]. Mascot takes a probability based approach to match peptide sequences from a database to tandem mass spectra, but as this is a commercial product, the exact

methodology behind the software is not known [Perkins et al., 1999]. Andromeda compares identified peptides to a library of hypothetical peptides, taking fixed and variable modifications into account. The probability score is calculated by counting the number of matches between the theoretical peptide fragment masses and the observed fragment ion spectral peaks. The higher the number of matches, the less likely the fragment has occurred by chance and hence a high probability score will be conferred to that peptide. Higher sequence coverage of a protein by the peptide identifications will lead to greater confidence in the true identification of a protein [Cox et al., 2011]. In addition, to determine the false discovery rate (FDR) of these identifications, the database search is also performed on a decoy database containing reversed protein sequences. The scoring thresholds are then adjusted to achieve a particular FDR, in most cases 1% at both the protein and the peptide level. MaxQuant software, which contains Andromeda as a search engine, provides quantitative data on peptides and their cognate proteins [Cox and Mann, 2008].

Top-down vs bottom up proteomics

The approach described above, whereby protein-containing samples are digested into peptides prior to mass spectrometric analysis, is referred to as a "bottom-up" approach. Trypsin is the most commonly used protease in mass spectrometry and it cleaves peptides on the carboxyl side of arginine and lysine residues. LysC is an additional agent which is becoming more frequently used, and this also cleaves peptides at the carboxyl side of lysine residues. Combination of tryspin and LysC digestion has been reported to improve sequence coverage by 8% [Ly et al., 2014].

Top-down proteomics refers to an approach where intact proteins are ionised directly and then analysed by MS. In theory, this approach would permit wider coverage of

protein sequences and PTMs but technical difficulties in this approach, not least resolution limits, mean it has so far lagged behind bottom-up proteomics [Kelleher, 2004; Catherman et al., 2014].

1.2.2 Sample fractionation prior to MS

To improve the depth of sequence coverage, with the aim of identifying a higher number of proteins within a sample, many sample fractionation techniques have been employed prior to MS analysis. Liquid chromatography has been widely used, relying on the differential separation of an analyte between a liquid mobile phase and a stationary phase. As discussed above, reverse-phase liquid chromatography uses a non-polar stationary phase and a polar mobile phase which allows more hydrophilic analytes to elute from a chromatography column first and more hydrophobic analytes last. Hydrophobic interaction liquid chromatography (HILIC) uses a polar stationary phase and non-polar mobile phase, with increasing water concentration, which causes hydrophobic analytes to elute first and hydrophilic analytes last [Alpert, 1990]. Ion exchange chromatography uses a salt gradient to differentially elute analytes on the basis of charge [Ritorto et al., 2013]. Separating proteins on the basis of size by SDS-PAGE has also been used prior to extraction and digestion as a non-LC based approach [Larance et al., 2012]. In addition, isoelectric focusing relies on differences in isoelectric point to separate different analytes [Cargile et al., 2004]. Fractionating cells into their different compartments, i.e. cytoplasm, membrane, soluble nuclear component, chromatin bound proteins etc. has also been applied [Boisvert et al., 2012; Larance et al., 2013]. All of these techniques generate multiple fractions from a single sample, which can then be introduced to the MS instrument sequentially with the aim of capturing more peptide identifications.

1.2.3 Protein quantification by mass spectrometry

The proteomic field has improved protein detection through advances in both sample preparation and instrument technology. However, as the proteome is dynamic, the focus of several groups has been to develop techniques to quantitate the relative abundance of proteins across multiple samples. Several approaches for protein quantitation by isotopic labelling exist (see [Tao and Aebersold, 2003] and [Goshe and Smith, 2003] for reviews). These techniques rely on the use of chemically identical, but mass-differentiated, stable isotopes, which can be used to label different sample groups. On mixing the samples and analysing by MS, the analyte from one sample represents a standard against which the chemically identical analyte from the other sample can be compared. Some of the most commonly used techniques are described in more detail below.

The first technique developed was Isotope-coded affinity tags (ICAT) [Gygi et al., 1999a]. This technique involves the addition of a reagent comprising an affinity tag (biotin), an isotopically labelled linker (heavy or light deuterium) and a thiol specific reactive group, to a cellular lysate prior to trypsin digestion and affinity purification for biotin. The reagent binds to cysteine residues in proteins and is therefore limited to the analysis of peptides containing this residue, which reduces sequence coverage.

Isobaric tagging for relative and absolute quantitation (iTRAQ) involves the addition of a reporter tag to a sample after trypsin digestion, therefore at the peptide level [Ross et al., 2004]. The isobaric tags form an amide linkage to any amine residue, i.e. either N-terminal, or within a lysine side chain. Quantitation of peptides with this approach, unlike ICAT, is deduced from the relative concentration of reporter ions and therefore is calculated at the MS/MS level, rather than MS level. An advantage of this approach

over ICAT is that all peptides are potentially available for analysis thus providing greater sequence coverage. In addition, the availability of multiple different tags enables multiple different experimental parameters to be compared at one time. Alternatively, a selected number of experimental conditions can be compared and a different set of tags can be used to label synthetic peptides for use as reference proteins for absolute protein quantitation.

Stable isotope labelling by amino acids in cell culture (SILAC) was developed as an alternative approach in 2002 [Ong et al., 2002a] and the principles of this technique are depicted in Figure 1.13. This was one of the most effective approaches to label amino acids *in vivo* to enable relative protein quantitation. Cell culture medium depleted of a specific standard amino acid is supplemented with an isotopically labelled form of the same amino acid, which is then incorporated into proteins, usually with little or no effect on cell growth. Protein containing lysates incorporating “heavy” or “light” amino acids can be mixed immediately after harvesting, therefore eliminating experimental variability between the experimental and control groups in the subsequent digestion and sample preparation steps. Amino acids which have been isotopically labelled for the SILAC approach include leucine [Ong et al., 2002a], tyrosine [Ibarrola et al., 2004], arginine [Ong et al., 2002b] and lysine [Martinović et al., 2002]. This tag-less approach holds advantages over other quantitative approaches as it permits the analysis of the whole proteome and PTMs can also be studied. The SILAC approach has been developed to incorporate three isotopic labels thereby facilitating the comparison of a wider range of experimental scenarios. This triple SILAC approach [Blagoev et al., 2004] has been widely used in the Lamond laboratory for the study of protein turnover [Boisvert et al., 2012], turnover and subcellular localisation of protein

isoforms [Ahmad et al., 2012], proteome changes in response to drug treatment [Larance et al., 2012] and global subcellular characterisation of protein degradation [Larance et al., 2013]. The SILAC approach to quantitative proteomics has been extended by multiple groups to include nematodes [Larance et al., 2011], mice [Krüger et al.], yeast [Gruhler et al., 2005] and drosophila [Xu et al., 2012].

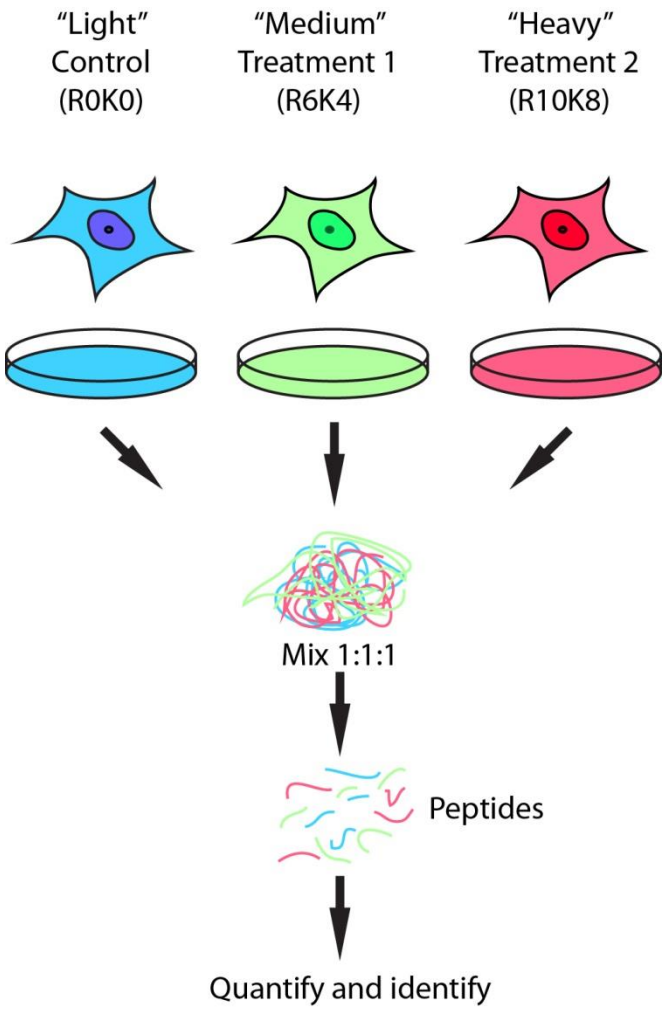


Figure 1.13. Schematic diagram of a triple SILAC experiment.

Schematic diagram of a triple SILAC experiment involving "light", "medium" and "heavy" labelled peptides.

1.3 Protein Interaction Analysis

1.3.1 The study of protein-protein interactions

The majority of proteins function as part of multiprotein complexes and not as isolated polypeptides. Simply generating a list of proteins is not biologically useful when trying to understand cellular machinery – all the protein interactions, and complexes they form must be delineated too. These protein complexes range from simple homodimers to large structures composed of many different polypeptides. Protein complexes vary in their size and shape from small globular dimers, such as 14-3-3 proteins, to large elongated filaments of variable length, such as microtubules. The wide variety of possible protein–protein interactions (PPI) within multiprotein complexes contributes to the diversity of functions that are involved in cellular processes and regulatory mechanisms. Several methods have been utilised over the years to study PPIs, some of which are described in the following paragraphs.

Yeast two-hybrid screen

Early efforts for high-throughput analysis of PPIs were predominantly undertaken in yeast and included the use of the yeast two-hybrid screen (Y2H). The screen is a molecular biology technique that has been used for many years to identify novel protein interaction partners. This approach uses a split transcription factor, one domain (the binding domain) fused to a bait protein in a plasmid, and the other (the activation domain) fused to another protein expressed most often as part of a "screening library" which may contain all the proteins expressed in an organism. If the bait and a target protein interact, the two transcription factor components are activated, and transcription of a downstream reporter ensues in the nucleus of yeast (Figure 1.14).

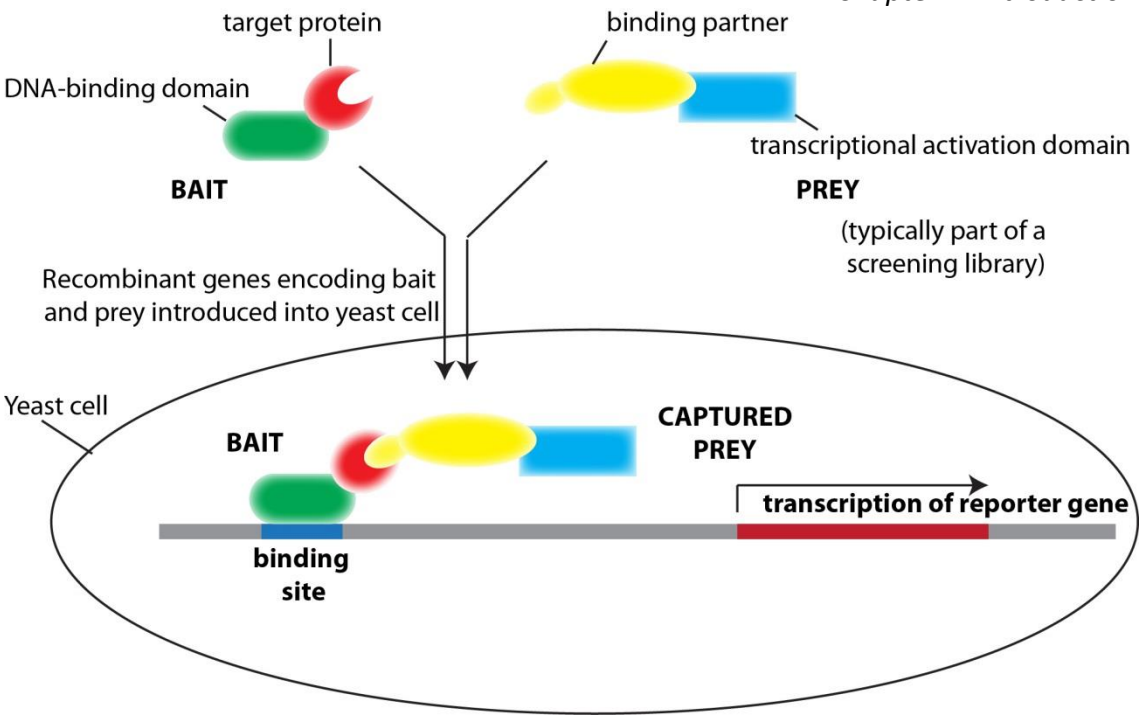


Figure 1.14. Schematic diagram of a yeast two-hybrid screen set-up.

Limitations of this approach are numerous, and include the inability of environmental factors to be taken into account on PPIs. In addition, the Y2H approach requires the cDNA library to be well annotated to improve the accuracy of results which is not always the case [Stelzl and Wanker, 2006]. False positive interactions are common, therefore validation of results by as many alternative approaches as possible is recommended [von Mering et al., 2002; Tikhmyanova et al., 2007]. The requirement that interacting proteins must translocate into the yeast nucleus demonstrates that this approach is not appropriate for the study of membrane proteins, cytosolic proteins, or proteins contained in other subcellular compartments [Bruckner et al., 2009]. Reports of bias against the detection of proteins involved in translation have also been associated with the Y2H screening method [von Mering et al., 2002]. As mentioned previously, post translational modifications have a significant role to play in PPI networks and some PTMs are not generated in yeast therefore a huge amount of

data regarding PPIs is not possible to detect by this approach [Bruckner et al., 2009]. Although, the Y2H screen can provide useful information in the study of PPIs, the binary nature of the interactions limit their usefulness for high-throughput analysis of large interaction networks.

Co-immunoprecipitation

Co-immunoprecipitation (co-IP) is perhaps one of the simplest and most widely used techniques to study PPIs. Antibodies raised against a target protein, or to a "tag", are added to a mixture of cellular proteins (Figure 1.15). An affinity reagent (e.g. IgG) linked to a solid bead matrix can then be used to immunoprecipitate the antibody and proteins interacting with the target protein, before eluting from the matrix. Interacting proteins can then be identified by immunoblotting. This technique requires the availability of highly specific antibodies, which are not always available and are expensive to create.

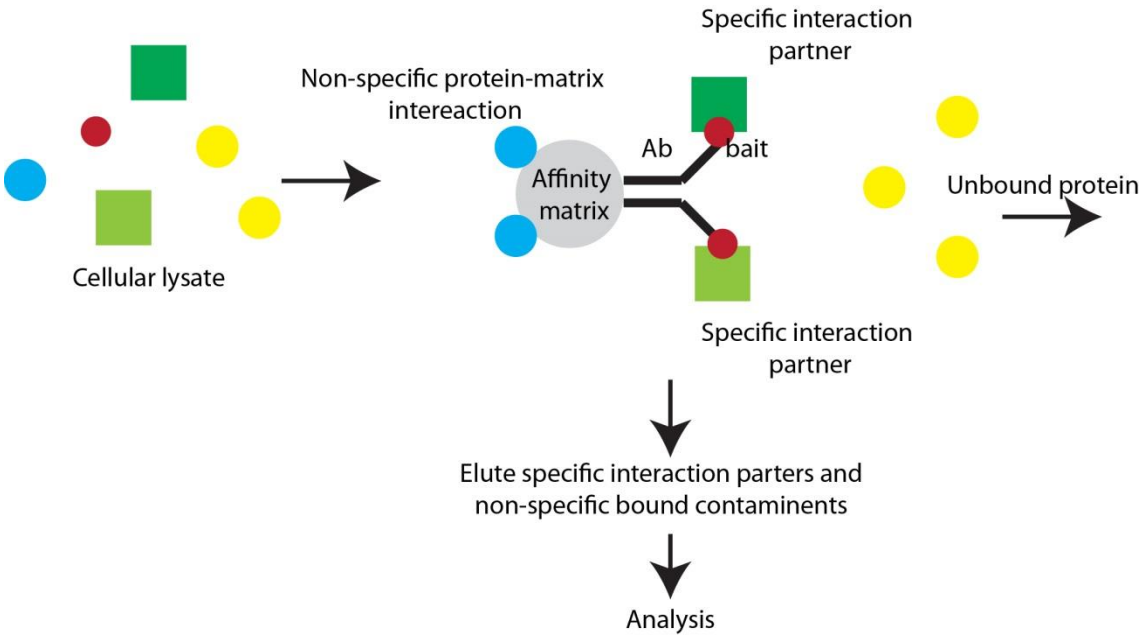


Figure 1.15. Schematic illustration of a typical immunoprecipitation analysis.

Protein affinity chromatography

Protein affinity chromatography involves the attachment of a bait protein to polymer beads which are then packed into a column. Cellular proteins are then passed through the column and interacting proteins will be trapped by the target-bead bait. Interacting proteins can then be eluted from the beads and identified by a variety of techniques.

Co-IP and affinity purifications have similar limitations in that in order to be pulled down, a PPI needs to be stable enough in the chosen experimental conditions. Therefore, either weakly associated, or transient, interactions may not be identified. In addition, either antibody binding sites, or parts of the protein involved in the covalent attachment to beads, may be involved in other PPIs, which would therefore not be identified in studies utilising this approach. In attempts to resolve this, epitope-"tags" (e.g. green fluorescent protein, (GFP)) can be expressed at either the C- or N-termini of proteins through recombinant DNA techniques and immunoprecipitation can be performed using antibodies to the protein tag. Alternatively, recombinant DNA techniques can be used to fuse a small enzyme, such as glutathione S-transferase (GST), to the bait protein, prior to binding the resulting fusion protein to glutathione beads. The beads can then be used to pull down interacting proteins from a cellular protein mixture for further analysis. Expressed, tagged proteins are often under the control of an inducible promoter and as a result, altering the stoichiometry of the protein of interest may affect the protein complexes it interacts with.

Tandem affinity purification

Tandem affinity purification relies on the same principle as epitope tagging, except two successive tags are used instead of one. The TAP tag comprises a ProtA IgG binding-

unit and calmodulin binding protein (CBP), specifically selected as they enable efficient pull down when expressed at low/near endogenous protein levels in cells [Rigaut et al., 1999]. The tag is fused to a clone of interest, and then expressed in a cell. After cell lysis the TAP tag is pulled down using IgG-trap beads as bait. The bound proteins are then eluted from the beads by cleaving a specific part of the TAP tag. The second tag remains bound to the protein of interest. Therefore, a second pull down for this tag can be performed before eluting the immunoprecipitated proteins. The rationale for this double tagging approach is to reduce non-specific background on the basis that non-specific protein interactions are reduced even more-so than with single tag methods.

A major limitation of all these tag-based approaches is that tagging a protein can alter protein structure and as a result may preclude the formation of some interactions [Goel et al., 2000; Rumlová et al., 2001]. Additionally, some non-specific interactions may occur as a result of proteins binding either beads, or tags, rather than the bait protein itself. Therefore, adequate control IPs must be run in parallel to identify specific interaction partners.

Such antibody-antigen methods for studying PPIs require a degree of prior knowledge about potential interaction partners to guide further investigations. The advent of mass spectrometry as a technique for studying proteins, however, changed this immensely by providing a high-throughput, unbiased approach for the identification of truly unknown PPIs.

Size exclusion chromatography

Size-exclusion chromatography (SEC) is a well-established technique used to separate proteins and protein complexes in solution on the basis of their shape/size (rotational

cross-section) [Wheaton and Bauman, 1953]. SEC has been extensively used as an intermediate step in conventional multistep biochemical protein purification strategies. In the study of protein interactions, researchers have traditionally targeted SEC fractions for specific proteins of interest by immunoblotting, but its use for monitoring protein complexes has been limited [Olinares et al., 2010].

The apparatus for SEC involves a column packed with polymer beads, with each bead designed to contain multiple pores of a defined size. A protein containing sample is pumped through the column at a continuous rate and as the proteins are contained in a liquid, they are free to rotate. Proteins vary widely in structure, from linear to globular, therefore rotation in three dimensions increases their cross-sectional size and it is this size that dictates passage through pores, rather than molecular weight. Depending on pore size, either large proteins, or large protein complexes, will not fit into any of the pores and will therefore migrate through the column rapidly and elute in the first fractions. The earliest fractions where large proteins and complexes cannot be separated at all are termed the “void”. Smaller proteins and complexes pass into the pores and as a result, their migratory progress through the column is slower in comparison to larger proteins, essentially because they must move through a larger volume. As a result, the proteins in an analyte will be separated in the column on the basis of rotational cross-section size, with the largest proteins or complexes eluting early and smaller proteins/complexes eluting last. The eluate produced by the column over time is collected in multiple fractions. If proteins interact they would be expected to co-elute in the same fractions. A schematic diagram of this process is described in Figure 1.16.

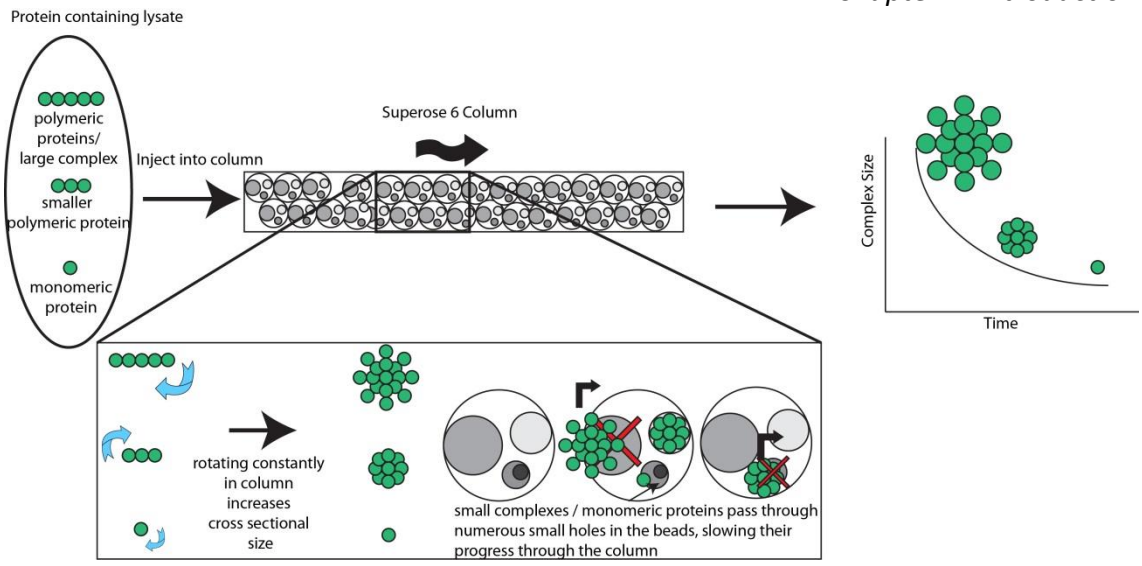


Figure 1.16. Schematic diagram summarising size exclusion chromatography.

Native proteins and protein complexes are separated on the basis of rotational cross-section size. Larger proteins/complexes elute in the earliest fractions and smaller proteins/complexes elute later from the column.

Protein microarrays

Additional *in vitro* approaches for detecting PPIs include protein microarrays. This is a high-throughput method which enables the identification of protein interactions and function. In principle, the technique involves binding a protein (or several proteins) to a support medium, then probing with fluorescently labelled molecules. Any reaction between the probe and immobilised protein emits a fluorescent signal which is then read by a laser scanner [MacBeath and Schreiber, 2000]. This approach has the advantage that it can be carried out on a large scale in an automated process. However, the preparation of proteins to analyse can be a limiting factor as it requires high-quality and comprehensive expression libraries and array product methods that yield a large number of functionally active proteins [Droit et al., 2005]. In addition, non-specific interactions and cross-reactivity can be problematic in this technique. This approach also provides only a binary output (i.e. either proteins interact or they do

not) and does not readily enable the quantitation of protein abundance, or the measurement of PTM states [Wolf-Yadlin et al., 2009].

X-ray crystallography

X-ray crystallography enables the visualisation of protein structures at an atomic level and as a result, it can show how proteins interact with other molecules [Garman, 2014]. This technique can be applied to study the conformational change undertaken by a protein in response to a modification. Applications of this approach rely on the ability to crystallise a protein of interest and it is not suitable for large scale analyses of protein interactions in a system-wide study.

1.3.2 LC-MS/MS applied to protein complex analysis

The affinity purification tagging approach is now widely used in combination with mass spectrometry and is advantageous in that many different complexes can be compared using an identical antibody, or other affinity-purification method, targeted to the tag on the bait (for examples, see Refs. [Ho et al., 2002; Gavin et al., 2006; Ewing et al., 2007]; for reviews, see [Chang, 2006; Collins and Choudhary, 2008] and [Gingras et al., 2007]). In contrast, it is harder to compare directly the results from immunoprecipitation of different endogenous protein complexes by MS because each specific antibody that is used has different affinities and properties. Malovannaya *et al.* have developed a high-throughput approach for the identification of endogenous protein complexes through cross referencing the results of over 1,000 individual endogenous immunoprecipitations combined with LC-MS/MS [Malovannaya et al., 2010]. By utilising this approach, the same group were able to study the endogenous coregulator complexome and identify previously unreported proteins in this network [Malovannaya et al., 2011]. Although endogenous IPs avoid the problems associated

with an epitope tag, and enable the experiment to be performed in near endogenous conditions, the sheer cost involved in large scale antibody production is beyond the scope of most laboratories. Attempts have been made to produce antibodies on a large scale to explore the human proteome by tissue analysis [Nilsson et al., 2005] but the suitability of these antibodies for interaction analysis would need to be tested further.

To determine whether co-purifying proteins are either *bona fide* interaction partners, or non-specific contaminants, quantitative proteomic approaches based on the SILAC approach have been combined with affinity purification [Trinkle-Mulcahy et al., 2006; Trinkle-Mulcahy et al., 2008; Hubner et al., 2010]. Additional data analysis procedures, including the use of a "super experiment" database that predicts the likelihood of non-specific protein interactions based on the frequency with which any given protein is co-purified across many separate experiments, can also help to define the composition of protein complexes [Boulon et al., 2010]. The combination of AP-MS with quantitative proteomics can also be used to study the dynamic changes in protein interactions such as those that change in response to inhibition of NEDDylation [Larance et al., 2012]. Nonetheless, affinity purification strategies have a limited ability to distinguish multiple related complexes that may differ with respect to isoforms and PTMs. They are also costly and difficult to implement for large-scale studies to survey cellular complexes, and thus not well suited to study variations in complexes under different cellular growth conditions and responses.

For system-wide studies of the composition and dynamics of protein complexes, alternative methods, in addition to immune-affinity purification, are required for convenient separation, characterization, and comparison of cellular protein complexes.

To address this, a number of studies have utilized various forms of differential centrifugation, column chromatography or native gel electrophoresis, to isolate complexes or organelles of interest, in combination with mass-spectrometry-based proteomics.

The use of ultracentrifugation permits the fractionation of protein complexes on the basis of differential sedimentation rate, which is predominantly determined by molecular weight. This approach combined with a sucrose density gradient has been used for the isolation of ribosomal subunits prior to MS analysis in Arabidopsis [Chang et al., 2005] and yeast [Link et al., 1999]. Although this can be an easily applied technique, disadvantages include the contamination of the isolated fractions by proteins with a similar sedimentation co-efficient [Chang, 2006].

Protein complexes have also been separated using blue native polyacrylamide gel electrophoresis (BN-PAGE) [Camacho-Carvajal et al., 2004; Heide et al., 2012].This approach involves loading a protein containing analyte into a gel matrix which then separates complexes on the basis of their molecular weight, in a range from 10kDa to 10MDa. In comparison to some gel filtration columns and centrifugation approaches, the resolution obtained by BN-PAGE is high. However, disadvantages of this approach include the limited amount of protein that can be loaded into the gel in one run. In addition, isolation of the complexes involves gel cutting and in-gel digestion which introduces error and variability when comparing biological and/or technical replicates.

Ion-exchange chromatography (IEX) was utilised by Havugimana *et al.* in what was one of the first attempts to separate all identifiable protein complexes within the nuclear and cytoplasmic extracts of two cell lines (HeLa and HEK293 cells), using a continuous in-solution approach [Havugimana et al., 2012]. The novelty in this study lay in the use

of multiple column combinations to validate and score protein co-elution patterns, increasing the probability that co-eluting fractions were interacting. A disadvantage of this approach, however, is that the salt concentrations in the buffers for ion exchange can perturb some PPIs, thereby preventing their detection.

Size exclusion chromatography is a different chromatographic approach, which has been utilised in combination with LC-MS/MS to study the native protein complexes in plant chloroplasts [Olinares et al., 2010] and large cytosolic complexes in mammalian cells [Kristensen et al., 2012]. The advantage of the SEC system for protein complex analysis is that almost any buffer can be used. As a result, the lysis of cells in a buffer containing physiological levels of salt will preserve many native protein complexes and separating these by SEC permits further analysis of the composition of native protein complexes.

1.3.4 Bioinformatics in protein complex analysis

Interaction Prediction

One significant difficulty raised by large-scale protein complex analyses is data management. As increasing numbers of large-scale experiments are being performed, there is a greater need for tools to analyse, visualise, validate and manage the data.

In the previous sections I have described different *in vitro* techniques used to study PPIs. Several *in silico* approaches (i.e. computational analysis) have also been employed to predict protein interaction sites on the basis of protein sequences [Ofran and Rost, 2003; Hosur et al., 2011], evolutionary preservation of protein interaction sites [Lichtarge and Sowa, 2002; Glaser et al., 2003], gene expression [Grigoriev, 2001] and protein structure [Aloy and Russell, 2002, 2003; Singh et al., 2010]. These

approaches have been reviewed previously [Ethier et al., 2006; Rao et al., 2014]. As more data emerges about different proteins, these databases will expand and will act as a valuable guide to potential protein interactions. However, many of the predicted interactions remain to be proven experimentally and it remains to be established by experiment under which circumstances (e.g. cell type, compartment, cell cycle phase) these interactions occur.

Protein interaction databases

As I have illustrated in the previous sections, no large scale PPI experiment is immune to false positive interactions. Although there are discrepancies in protein interactions derived by different experimental techniques, several databases have been developed that collate interaction data to increase confidence in results [Droit et al., 2005]. Some of these are summarised in Table 1.2. As many medium to large scale studies develop their own databases for protein interaction analyses, this list is not exhaustive.

Database Name	Source link	Data source
DIP	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	Collated experimentally derived interactions
BioGrid	http://thebiogrid.org	Collated experimentally derived interactions
HitPredict	http://mintdb.hgc.jp/htp/	IntAct, BioGrid, HPRD
MINT	http://mint.bio.uniroma2.it/mint/	Collated experimentally derived interactions
IntAct	http://ebi.ac.uk/intact	Collated experimentally derived interactions
APID	http://bioinfow.dep.usal.es/apid/index.htr	BIND, BioGrid, DIP, HDRP, IntAct, MINT
BIND	http://bind.ca	Collated experimentally derived interactions
STRING	http://string-db.org	Collated experimentally derived interactions, text mining, interologs, gene expression
MPPI	http://mips.helmholtz-muenchen.de/proj/	collated experimentally derived interactions
HPRD	http://hprd.org	collated experimentally derived interactions

Table 1.2. Representative selection of publicly available protein interaction databases

STRING is a visualisation tool which allows the user to identify their protein(s) of interest and its interactions with other proteins are depicted by a ball and string model [Snel et al., 2000; Szklarczyk et al., 2010]. This database collates both experimental and predictive data from published databases. However, selection bias in the original

experiments may mean that interaction data are only available for previously annotated proteins. These databases, however, focus on protein interaction partners and do not easily discriminate separate protein complexes.

The CORUM database, compiled using a variety of information from the literature describing protein interactions and assemblies, provides the largest public dataset of protein complexes [Ruepp et al., 2010]. CORUM contains information relating to ~1,970 protein complexes identified in human cells. However, these proteins are formed from proteins encoded by only ~16% of the known protein coding genes, indicating that many forms of protein complexes remain to be identified and characterised.

Computational mathematical analysis of protein complex networks

Protein complex identification is the next step in determining protein function beyond simple PPIs. Lappe and Holm describe a method to unravel protein interaction networks through the collective analysis of multiple pull down experiments combined with mathematical modelling based on proteins (nodes) and their interactions (edges) [Lappe and Holm, 2004]. They used analysis of the yeast proteome to test their theory and on the basis of these results, propose that 90% of the human interactome could be analysed from the collective analysis of 10,000 pull down experiments [Lappe and Holm, 2004]. Malovannaya *et al.* combined large scale endogenous IP analyses (>1,000 endogenous pull down analyses) with a "near neighbour network" analysis to identify reciprocal protein-protein interactions. Clustering analysis of the top IP results co-immunoprecipitating with the protein of interest, enabled this group to identify possible protein complex components [Malovannaya et al., 2010]. The same group used this approach to identify previously unknown protein interactions in the human

coregulator complexome [Malovannaya et al., 2011]. Limitations of this approach stem from the high cost and impracticality of performing such large scale endogenous immunoprecipitations for most laboratories.

Hierarchical clustering has been used in combination with size exclusion chromatography and mass spectrometry to analyse megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* [Olinares et al., 2010]. This approach clusters proteins on the basis of the similarity of their elution profiles, the assumption being that proteins with similar profiles form part of the same protein complex. Olinares *et al.* utilised this approach to separate chloroplast ribosomes in different states of assembly, and enabled the identification of plastid homologues of prokaryotic ribosome assembly factors as well as proteins involved in co-translational modifications, targeting and folding. Kristensen *et al.* [Kristensen et al., 2012] also combined SEC and quantitative mass spectrometry with hierarchical clustering to study the effect on the interactome of EGF stimulation. Havugimana *et al.* [Havugimana et al., 2012] performed an analysis of human protein complexes utilising hierarchical clustering of chromatography and sucrose gradient data, combined with predicted protein interaction from evolutionary data, mRNA co-expression and domain co-occurrence. This dataset provides information regarding previously unreported complexes with specific functional, evolutionary and disease-related biological attributes.

Although current methods are unable to predict interactions and protein complex components with 100% accuracy, the development of computational methods from a range of experimental techniques will provide an invaluable resource for guiding future laboratory based experiments.

1.4 Summary

Studying cell biology at the protein level with the current techniques available provides a very detailed analysis of cell function, which would not be possible by studying the genetic code or mRNA at the transcriptional level. Large scale proteomic analysis has enabled the detailed study of protein localisation at the subcellular level, including the localisation of distinct protein isoforms, demonstrating that different isoforms may have different functions [Ahmad et al., 2012]. The study of protein turnover rates demonstrated that regulation of protein expression can vary between subcellular compartments without necessarily altering overall protein levels [Larance et al., 2013]. This suggests that regulation of protein expression at the subcellular level is important in regulating cell function. The proteomics approaches utilised in current studies has also enabled the most in-depth analysis to date of the proteome as it changes throughout the cell cycle, demonstrating again that RNA transcription does not accurately reflect protein expression throughout this process [Ly et al., 2014]. In addition to all of the above, the mass spectrometry approach also enables the study of protein abundance through quantitative analysis [Goshe and Smith, 2003; Tao and Aebersold, 2003], and the study of PTMs on a large, high-throughput scale [Olsen et al., 2006; Choudhary et al., 2009; Wagner et al., 2011; Tammsalu et al., 2014]. Because all of these factors are involved in protein interactions, it is important to utilise a technique that enables all of these factors to be studied to facilitate a greater understanding of cell function.

Chapter 2: Materials and methods

2.1 General materials

PhosStop protease inhibitor cocktail tablets and Complete EDTA-free protease inhibitor tablets were obtained from Roche. NuPAGE Bis-Tris gels, NuPAGE Tris-acetate gels and NuPAGE LDS sample buffer were from Life Technologies and TCEP was from Thermo Fisher Scientific (77720). MLN4924 was a generous gift from Sir Philip Cohen. All chemicals were obtained from Sigma unless otherwise specified. Testes from C57BL/6 mice were generously donated by members of the Cantrell Laboratory (University of Dundee). Human testicular tissues were a generous gift form the Department of Pathology, Edinburgh Royal Infirmary.

2.2 Antibodies

The antibodies utilised in this thesis are listed in Table 2.1.

Antibody	Company	Catalogue number	Source	Utility
MRFAP	ProteinTech Group	11639-1-AP	Rabbit	WB, IF
MORF4L1	Santa Cruz	sc-26528-R	Rabbit	WB, IP, IF
VprBP	Bethyl	A310-887A	Rabbit	WB, IP
MRGBP	Sigma/Atlas antibodies	HPA017012	Rabbit	WB
MRGBP	Tebu-bio	H00055257-B01P	Mouse	IF
Lamin A/C	Cell Signalling technololgy	#4777	Mouse	WB
GFP	Roche	11814460001	Mouse	WB
Histone H3 (D1H2)	Cell Signalling Technology	4499	Rabbit	WB
CUL3	Cell Signalling Technology	#2759	Rabbit	WB
eIF2α	Cell Signalling Technology	#2103	Mouse	WB
Phospho-eIF2α	Cell Signalling Technology	#3398	Rabbit	WB
Fibrillarin	Abcam	ab18380	Mouse	IF
Tubulin/DM1A	Sigma	T6199	Mouse	WB, IF
GAPDH	Abcam	ab8245	Mouse	WB
NEDD8	Cell Signalling Technology	#2745	Rabbit	WB
PAF	Abcam	ab56773	Mouse	WB
Beta-actin	Abcam	ab8227	Rabbit	WB
Alexa Fluor 488	Invitrogen	A21202	Mouse	IF
Alexa Fluor 594	Invitrogen	A21207	Rabbit	IF
HRP conjugated antibody	Thermo	PA1-84824	Mouse	WB
HRP conjugated antibody	Thermo	31460	Rabbit	WB
Alexa Fluor 680	Invitrogen	A21058	Mouse	WB
Anti-IgG IRDye 800	Tebu-bio	611-132-003	Rabbit	WB

Table 2.1. Antibodies used for experiments described within this thesis.

2.3 General solutions

TBS

150mM NaCl

10mM Tris pH 7.4

PBS

137mM NaCl

2.7mM KCl

10mM Na_2HPO_4

1.8mM KH_2PO_4

TBST

150mM NaCl

10mM Tris pH 7.4

0.1% Tween 20

2.4 Tissue culture

2.4.1 Cell lines and solutions

- U2OS, HaCaT, Tera 1 and U138 cells were purchased from the American Type Culture Collection (ATCC).
- The wild type LAP1-MRFAP1 and LAP1-7KR MRFAP1 U2OS cell lines were a generous gift from Dr Mark Larance, University of Dundee.

Denaturing lysis buffer:

2% SDS

10mM HEPES pH 7.4

1mM EDTA

250mM Sucrose

1x Complete EDTA-free protease inhibitor tablet per 50ml lysis buffer and 1 x PhosStop tablet per 10ml buffer.

2.4.2 Tissue culture conditions

U2OS, U138 and HaCaT cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies) supplemented with 10% foetal calf serum (FCS, Life Technologies), 100U/L penicillin, and 100µg/L streptomycin.

U2OS cells expressing LAP1-tagged MRFAP1 and the LAP1-7KR MRFAP1 mutants were grown in DMEM supplemented with 10% FCS, 100U/L penicillin, and 100µg/L streptomycin, 150µg/ml hygromycin B and 15µg/ml blasticidine HCl.

Tera-1 cells were cultured in McCoy's 5a medium (Life Technologies) supplemented with 20% FCS, 100U/L penicillin, and 100µg/L streptomycin.

All cell lines were grown at 37 °C in 5% CO_2 and passaged at ~80% confluence.

2.4.3 SILAC Labelling

For SILAC labelling of U2OS cells, arginine- and lysine-free Dulbecco's modified Eagle's medium (Life Technologies) was used and supplemented with stable isotope-labelled arginine (Cambridge Isotopes) to a final concentration of 42μg/ml and lysine (Cambridge Isotopes) to a final concentration of 72μg/ml, in addition to dialysed FCS (Life Technologies). The final concentrations of arginine and lysine used in the SILAC medium were half of those contained in standard medium to suppress arginine to proline conversion [Ong et al., 2002a]. The final medium was filtered through a 0.2μm filter. Cells were passaged three times (1:6 split) before commencing an experiment, to ensure adequate labelling of proteins was obtained.

2.4.4 Cell lysis

Unless otherwise stated, cells were harvested at ~80% confluence. Cells were washed in PBS (10-20ml depending on whether a 10 or 15cm dish was used) then lysed in 200-500μl of denaturing lysis buffer, pre-warmed to 37^{o}C. The lysates were heated to 65^{o}C for 10 minutes then sonicated for 3 x 10 second bursts at 10% power (Branson Digital Sonifier 250). The lysate was centrifuged at 17,000g for 10 minutes using a Heraeus Pico17 centrifuge (Thermo Fisher Scientific) and the supernatant was transferred to a clean Eppendorf tube. BCA quantitation assay was performed on the supernatant as per section 2.16. An aliquot of lysate was made up to 1mg/ml in 1X LDS sample buffer with 25mM TCEP for immunoblotting.

2.5 SiRNA knockdown

2.5.1 Materials

- ON-*TARGETplus* Set of 4 Upgrade siRNA for MRFAP1, MRGBP and MORF4L1 were obtained from Dharmacon (catalogue numbers LU-015093-01-0002, LU-006381-01-0002, LU-006379-01-0002 respectively)
- Lamin A/C siRNA targeting the sequence 5'-CUGGACUUCCAGAAGAACA-3' was obtained from Dharmacon.
- Jumble siRNA targeting the sequence 5'-CAGUCGCGUUUGCGACUGG-3' was obtained from MWG.
- 5X siRNA buffer (Dharmacon (B-002000-UB-100) – 300mM KCl, 30mM HEPES-pH 7.5, 1.0mM $MgCl_2$). Diluted to 1X with sterile H_20.
- Lipofectamine RNAiMAX (Invitrogen, 13778)

2.5.2 siRNA knockdown for immunoblotting

siRNA stocks of 100μM and 10μM were prepared using 1X siRNA buffer and stored at -20^oC.

U2OS cells (1 x 10cm dish per knock-down) were grown to ~40% confluence. Cells were washed (5ml per dish) in Opti-MEM (Life Technologies) and 10ml Opti-MEM was added to cells. 1ml of Opti-MEM was mixed with 45μl of 10 μM siRNA stock in a sterile Eppendorf tube. 1ml of Opti-MEM was mixed with 30μl Lipofectamine RNAiMAX (per siRNA knock-down). The RNAiMAX and siRNA dilutions were mixed in a 15ml tube and incubated for 15 minutes at room temperature. The mixture was then added to cells dropwise, and gently rocked back and forth to mix evenly. The cells were incubated for

48 hours at 37^{o}C, 5% CO_2. Cells were then harvested as described in section 2.4.4 and immunoblotted following SDS-PAGE as described in section 2.6.

2.5.3 siRNA knockdown for immunofluorescence

Cells were grown on 13mm glass coverslips in 24-well plates to 50% confluence. Medium was removed and the cells were washed in 0.5ml Opti-MEM then a fresh 0.5ml Opti-MEM was added to each well. 107µl Opti-MEM was mixed with 3.21µl 10mM siRNA stock in a 1.5ml Eppendorf. 107µl Opti-MEM was mixed with 2.14µl RNAiMAX in a second Eppendorf tube, and then the two mixtures were combined, mixed and left to incubate for 15 minutes at room temperature. The mixture was then added to cells dropwise, and gently rocked back and forth to mix evenly. The cells were incubated for 48 hours at 37^{o}C, 5% CO_2. MLN4924, 1µM final concentration, or 10µl DMSO control was added to cells 18 hours prior to fixation.

2.5.4 Statistical analysis of siRNA knockdown immunoblots

Immunoblot band intensity was calculated for each protein using Macbiophotonics ImageJ software. The signal intensity for each protein in a replicate was standardised against the tubulin signal to correct for errors in total sample loading per lane. The corrected signal intensity was then normalised against the "jumbled siRNA" signal as an internal control for each blot. Standard deviations were calculated from the results across three biological replicates and a pair-wise comparison was analysed between the final normalised results for each siRNA knockdown combination. The p-values from this comparison were then corrected for the false discovery rate by using the Benjamini and Hochberg method by Dr Nick Schurch of the Data Analysis Group, (GRE, University of Dundee) [Benjamini and Hochberg, 1995].

2.6 SDS-PAGE and immunoblotting

2.6.1 Solutions

Blocking buffer:

3% non-fat dried milk in TBS with 0.1% Tween 20

Transfer buffer:

1x NuPAGE transfer buffer (Life Technologies) supplemented with either 10% or 20% methanol.

Running buffer:

MES or MOPS (Life Technologies)

Gel Fixation buffer:	**Gel washing buffer:**
50% methanol	10% methanol
10% acetic acid	7% acetic acid
Milli-Q H_2O	Milli-Q H_2O

2.6.2 SDS-PAGE separation

Protein samples were dissolved in 1X LDS sample buffer supplemented with 25mM TCEP. Equal amounts of protein were loaded for SDS-PAGE with 10μg/lane, except for immunoprecipitation eluates, where 10% of the elution volume was loaded, and for combined SEC fractions, where protein concentration varied but did not exceed 10μg per lane. SDS-PAGE was performed using either 3-8% Tris-acetate NuPAGE gels, or 4–12% (w/v) Bis-Tris NuPAGE gels, using either MES (for low molecular weight proteins), or MOPS (higher molecular weight proteins), running buffer at 200V for either 45 minutes (Bis-Tris gels), or 1 hour 10 minutes (Tris-Acetate gels).

2.6.3 Immunoblotting

For Western blotting, separated proteins were electrophoretically transferred to an iBlot nitrocellulose membrane (Life Technologies, IB7010-01). Alternatively, proteins were transferred to an Odyssey nitrocellulose membrane with NuPAGE transfer buffer for 1 hour 30 minutes at 35V. Membranes were then blocked with 3% non-fat skim milk in 0.1% Tween 20 in TBS (TBST) for 1 hour at room temperature with agitation. Membranes were then incubated with primary antibody (Table 2.1) (1:1,000 to 1:5,000) in 3% non-fat milk in TBST overnight at 4°C. After incubation, the membranes were washed three times in TBST and incubated with either horseradish peroxidase-labelled (1:5,000), or Alexa Fluor 680/IRDye 800-labeled, secondary antibodies (Table 2.1) (1:15,000) in 3% non-fat skim milk in TBST. Proteins were visualized using Immobilon chemiluminescent substrate (Millipore) (5ml per membrane) and imaged either with a cooled CCD camera (LAS 4000, Fuji), for horseradish peroxidase-labelled secondary antibodies, or with a Li-Cor Odyssey CLx imager for Alexa Fluor 680/IrDye 800-labeled secondary antibodies. Western blots were viewed using either Macbiophotonics ImageJ software, or Image Studio Version 2.0 software, when the Li-Cor system was used.

2.6.4 SYPRO Ruby protein gel staining

SYPRO Ruby protein gel staining (Invitrogen, S-12000) was performed on SDS-PAGE gels run as above. The gels were fixed twice in 100ml gel fixation buffer for 15 minutes at room temperature with agitation. The gel was stained with 50ml SYPRO Ruby protein gel staining solution for 2 hours at room temperature, with agitation and protected from light. The gel was washed twice in gel washing buffer for 15 minutes at room temperature with agitation, and protected from light. The gel was subsequently

washed with Milli-Q water for at least 15 minutes and then visualised using the FLA 5100 imager (Fuji).

2.6.5 InstantBlue protein staining

SDS-PAGE gels run as described above were immersed in 20ml InstantBlue staining solution (Expedeon, ISB1L) for 1 hour at room temperature. The gel was then washed in MilliQ H_2O before being scanned by the Li-Cor Odyssey CLx instrument.

2.7 SEC: Non-crosslinked samples

2.7.1 Solutions and buffers

Native cell lysis buffer:

PBS, pH 7.2 (Gibco, 20012-019)

EDTA free complete protease inhibitor (1 tablet/50ml)

PhosStop (1 tablet/10ml)

Denaturing cell lysis buffer:

PBS containing EDTA free complete protease inhibitors and PhosStop (as above)

4% SDS

25mM TCEP

50mM N-ethylmaleimide

Non-denaturing SEC buffer:

PBS pH 7.2 (Gibco, 20012-019)

Denaturing SEC buffer:

0.2% SDS

100mM NaCl

10mM $NaPO_4$

2.7.2 Cell lysis

Ten 15-cm dishes (80% confluent) of U2OS cells were scraped, on ice, in 500µl of ice-cold cell lysis buffer. Total cell lysates were sonicated for 10 seconds, three times in total, at 10% power (Branson Digital Sonifier 250) at 4^{o}C and then centrifuged at 17,000g (Heraeus Pico17 centrifuge, Thermo Fisher Scientific) for 10 minutes at 4^{o}C. Samples were filtered through 0.45µm Ultrafree-MC centrifugal filter units (Millipore). Bradford assays were performed on the filtrates for protein quantitation (see section 2.18).

2.7.3 Tissue lysis

Both testes from wild type C57/BL6 mice, sacrificed at either 48, or 160 days, were washed three times in 10ml of ice cold PBS. Testes were lysed in 500µl of ice cold native cell lysis buffer and homogenised for 1 minute on setting 4 and 10 seconds on setting 6 using an IKA T8.01 S2 mixer, on ice. Lysates were sonicated for 3 x 10 seconds at 10% power, on ice (Branson Digital Sonifier 250). The lysates were then centrifuged at 17,000g for 10 minutes at 4^{o}C (Heraeus Pico17 centrifuge). The supernatant was filtered through a 0.45µm Ultrafree-MC centrifugal filter unit (Millipore). A Bradford assay was performed on the filtrate for protein quantitation (see section 2.18).

2.7.4 SEC, trypsin digestion and peptide clean-up

Using a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific), prepared lysates were injected (200µl per injection) onto a Superose 6 10/300GL column (23.5ml column volume) (GE Life Sciences) equilibrated with PBS (pH 7.2). The flow rate was 0.2 ml min^{-1}, and 40 x 200µl fractions were collected using a low protein-binding 96-deep-well Eppendorf plate. Tris-HCl (1M pH 8.0) was added to each fraction to a final

concentration of 0.1M to adjust the pH to 8.0, and trypsin, diluted in 0.1M Tris-HCl, was added at ratios of 1:50 - 1:100. The fractions were then incubated for 18 hours at 37^{o}C. For peptide desalting, trifluoroacetic acid (TFA) was added to a 1% (v/v) final concentration and peptides were purified using a Sep-Pak tC18 96-well µ-elution plate. The µ-elution plates were conditioned with the addition of 200µl acetonitrile per well and vacuum was used to pull the liquid through the plates. The plates were equilibrated by the addition of 200µl of 0.1% TFA to each well, pulled through by vacuum manifold. The samples were then loaded in to the plate and pulled through by vacuum manifold. The wells were then washed in 800µl 0.1% TFA to remove salts, then 200µl 0.1% TFA to remove excess buffer and salts. Peptides were eluted in 200µl of 50% (v/v) acetonitrile/0.1% TFA, and then dried using a Savant SPD 2010 SpeedVac concentrator (Thermo Fisher Scientific), prior to resuspension in 5% (v/v) formic acid. Peptide concentrations were determined using the CBQCA assay as described in section 2.17.

2.7.5 Denaturing SEC

For the denaturing SEC, cells were lysed in denaturing cell lysis buffer as described above in section 2.7.1. The column was equilibrated with denaturing SEC buffer. The lysate was otherwise injected onto the Superose 6 column as described above in section 2.7.4. Consecutive elution fractions were combined, heated to 65°C for 10 minutes, subjected to chloroform methanol precipitation (section 2.15, [Wessel and Flügge, 1984]), and resuspended in 1X LDS sample buffer with 25 mm TCEP for immunoblotting.

2.7.6 SDS-PAGE and immunoblotting of SEC fractions

For immunoblotting of non-denatured samples, Bradford protein quantitation assays were performed on the fractions (section 2.18). 20% SDS was added to each fraction to a 2% final concentration, and fractions were heated to 65°C for 10 minutes. 100µl of consecutive fractions were combined, and chloroform methanol precipitation was performed (section 2.15,[Wessel and Flügge, 1984]). Protein was then resuspended in equal volumes of 1X LDS sample buffer, 25mM TCEP so that the maximum concentration in the most concentrated fraction was 1mg/ml, and heated to 65°C for 10 minutes. Combined fractions were analysed via EZQ quantitation assay (section 2.19). 10µl of each combined fraction was loaded per lane for SDS-PAGE as described in section 2.6.2.

BCA protein quantitation (section 2.16) was performed on denatured samples. Equal volumes (14µl) of consecutive samples were combined and made up to a maximum of 0.1mg/ml in 1X LDS/TCEP. 20µl of sample was loaded per lane for SDS-PAGE (section 2.6.2). SYPRO Ruby staining was performed as per section 2.6.4. SDS-PAGE and immunoblotting was performed as per section 2.6.

2.7.7 LC-MS/MS and analysis of spectra

CBQCA quantitation assay was performed on the peptides as described in section 2.17. From these results, the injection volume for the strongest fraction was calculated to ensure a maximum of 1µg of peptide would be injected. The injection volumes were kept constant for all fractions to ensure elution profiles could be calculated accurately. Using a Thermo Fisher Scientific Ultimate 3000 nanoHPLC system, a maximum of 1µg of peptides in 5% (v/v) formic acid (~10µl) was injected onto an Acclaim PepMap C18 nano-trap column (Thermo Fisher Scientific). After being washed with 2% (v/v)

acetonitrile, 0.1% (v/v) formic acid, peptides were resolved on a 150mm × 75µm Acclaim PepMap C18 reverse-phase analytical column over a gradient from 2% acetonitrile to 80% acetonitrile over 100 minutes with a flow rate of 300nl min^{-1}.

The peptides were ionized by nano-electrospray ionization at 1.2kV using a fused silica emitter with an internal diameter of 5µm from New Objective. Tandem mass spectrometry analysis was carried out on an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) using collision-induced dissociation fragmentation of precursor peptides and fragment ion measurement in the linear ion trap. Data-dependent acquisition was performed with fragments of the top ten most intense peptide ions, a (TOP10) method described previously [Haas et al., 2006]. The RAW data produced by the mass spectrometer were analysed using the quantitative proteomics software MaxQuant [Cox and Mann, 2008] (version 1.3.0.5). This version of MaxQuant includes an integrated search engine, Andromeda [Cox et al., 2011]. The database supplied to the search engine for peptide identifications was the human UniProt database (June, 2011) containing 109,824 entries. The mass tolerance was set at 6ppm for precursor ions, and the MS/MS mass tolerance was set at 0.5Da. The enzyme was set as trypsin with up to two missed cleavages. Deamidation of Asn and Gln, oxidation of Met, pyro-Glu (with N-term Gln), and phosphorylation (STY) were set as variable modifications. N-ethylmaleimide on Cys was searched as a fixed modification. Identification was set to a false discovery rate of 1%. To achieve reliable identifications, all peptides identified were accepted based on the criteria that the number of hits in the “forward” database was at least 100-fold greater than the number of “reverse” database hits, thus resulting in a false discovery rate of less than 1%. The output from

MaxQuant provided peptide level data as well as protein level data, grouped by protein isoforms.

2.7.8 Data analysis of native SEC fractions

The following statistical analysis of native SEC fractions was performed by Dr Yasmeen Ahmad. To create an elution profile for an individual protein isoform in each of the three replicates, the number of peptides observed in each fraction was summed to generate a count of peptides per fraction (spectral counts). The resulting spectral count profile minima and maxima were normalized within the limits of 0 and 1, respectively. The average of the three elution profiles was based on the normalized spectral peptide counts for each fraction. Further downstream data interpretation for this study was performed primarily using the R language (version 2.15.1). From the three biological replicates, it was required that a protein be identified in at least two replicates with a minimum of two peptides in each. Proteins labelled as either contaminants, or reverse hits, were removed from the analysis. These stringent criteria left a subset of 8,165 proteins. For each protein, 120 quantitative measurements were recorded across the three replicates and 40 SEC fractions. The first six fractions were deemed to contain complexes outside the resolution range of the column and thus were excluded from further data analysis.

To analyse reproducibility between replicates, the peptide count profile across the remaining 34 fractions for each replicate was correlated between replicates. A quality filter was applied to ensure that a minimum of two out of the three pairwise correlations had a positive correlation coefficient. The biological replicates were collapsed by averaging to obtain a resultant mean peptide count profile describing the behaviour of each protein across the 34 fractions. The standard deviation for the three

biological replicates per fraction was also calculated to provide standard error bars on the protein profile graphs. The protein count profiles across the 34 fractions were normalized using the maximum peptide count for that protein and then hierarchically clustered based on Euclidean distance measurement and a "complete" agglomeration method. The output of the clustering was presented in a heatmap using the RColorBrewer library.

The gene ontology analysis was carried out using the DAVID Functional Annotation Tool for Biological Processes [Huang et al., 2008]. The full *Homo sapiens* proteome, supplied by DAVID, was selected as a background list. When cross-analysing with the CORUM dataset, a set proportion of proteins in a complex were required to co-fractionate to provide sufficient evidence that a complex had been identified. To calculate this proportion, the equation max (0.3, 2/*n*) × 100 was used, meaning either 30% of the components, or 2 divided by the number of proteins in the complex, whichever was greater, multiplied by 100. This approach ensured that for small complexes (*i.e.* containing six or fewer components), more than simply one or two of the component proteins needed to co-fractionate before labelling the complex as identified.

2.8 SEC: Crosslinked samples

2.8.1 Solutions

Formaldehyde crosslinking solution:

16% Formaldehyde Solution (w/v), methanol-free ampules (Thermo Fisher Scientific; #28909).

Make up to the desired percentage for crosslinking (e.g. 1-8%) in PBS fresh each time.

Quench buffer:

50mM Glycine in 1X PBS

Additional 2.5M stock of glycine in PBS.

Cell lysis Buffer:	**Denaturing SEC buffer:**
4% SDS	0.2% SDS
100mM NaCl	100mM NaCl
10mM $NaPO_4$ pH 6.0	10mM $NaPO_4$ pH 6.0
25mM TCEP	passed through a 0.2μm filter
50mM NEM	

2.8.2 Crosslinking and cell lysis

1 x 15cm dish of U2OS cells (~80% confluent) washed twice in 20ml PBS. Crosslinking was performed by the addition of 20ml of formaldehyde crosslinking solution and incubated at room temperature for 10 minutes. The crosslinking solution was removed then the reaction was quenched by the addition of 20ml quench buffer for 5 minutes. A second quench step was performed with a fresh aliquot of quench buffer. Cells were washed again in 20ml PBS and drained well. Cells were lysed by scraping in the presence of 500μl of cell lysis buffer, supplemented with 10μl of 2.5M glycine stock. The lysate was transferred to a 1.5ml Eppendorf tube then sonicated for 10 seconds x 6 (separated by 10 second intervals) at 10% power (Branson Digital Sonifier 250). The lysates were then incubated in a water bath a 37^{o}C for 30 minutes to aid denaturation whilst maintaining crosslinks intact. Lysates were centrifuged at 17,000g for 10 minutes at room temperature (Heraeus Pico17 centrifuge, Thermo Fisher Scientific). The supernatant was filtered through an Ultrafree-MC 0.45μm filters (Millipore) at 12,000g for 5 minutes. EZQ protein quantitation assay was performed on the

supernatant as per section 2.19. 450µl of filtered supernatant was transferred to a glass vial with a v-shaped base for injection onto the SEC column.

2.8.3 SEC, protease digestion and peptide clean-up

Using a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific), prepared lysates were injected (200µl per injection) onto a BioBasic Sec 1000 column (Thermo Fisher Scientific) equilibrated with denaturing SEC buffer. This is a 7.8 x 300mm column with a volume of approximately 14ml. It contains 5µm particles with 1000Å pores. The flow rate was 0.5ml min^{-1}, and 48 x 125µl fractions were collected using a 96 well PCR-plate (Thermo Fisher Scientific). Eluates were heated to 95^{o}C for 30 minutes using an Eppendorf Mastercycler PCR machine when crosslink reversal was required. When cooled, tetraethylammonium borohydride (TEAB) (1 M, pH 8.5 (Sigma)) was added to each fraction to a final concentration of 0.1M to adjust the pH to 8.0-8.5. Lysyl Endopeptidase (LysC) (Wako, 125-02543) diluted in 0.1M TEAB was added at a ratio of 1:50 (µg LysC to µg of protein in sample) in a total volume of 50µl. The fractions were then incubated for 4 hours at 37°C. Trypsin was then added at 1:50 ratio to each well in a 50µl total volume, diluted in 0.1M TEAB. Fractions were then incubated for 18 hours at 37^{o}C. SDS removal was performed using 96-well detergent removal spin plates (Thermo Fisher Scientific). For peptide desalting, trifluoroacetic acid (TFA) was added to a 1% (v/v) final concentration and peptides were purified using a Sep-Pak tC18 96-well µ-elution plate as described in section 2.7.4.

2.8.4 SDS-PAGE and immunoblotting

For immunoblotting of cell lysates treated with and without formaldehyde, 28µl of representative fractions were taken from each experimental group, to which 10µl of 4X LDS sample buffer and 2µl of 0.5M TCEP were added. The start material (total cell

lysate) was made up to 1μg/ml in 1X LDS sample buffer supplemented with 25mM TCEP in MilliQ H_2O. All samples were heated to 95^{o}C to reverse crosslinks. 30μl of each sample was loaded onto a 10 well 4-12% Bis-Tris NuPAGE gel and run as described in section 2.6.2. SYPRO ruby staining was then performed as described in section 2.6.4.

To calculate the optimum concentration of formaldehyde for crosslinking, an EZQ assay was performed on the total cell lysates of crosslinked cells and the samples made up to 1mg/ml with 1X LDS in MilliQ H_2O. Samples were not heated further to prevent crosslink reversal/breakage. 10μl of each sample were loaded onto 4-12% Bis-Tris NuPAGE gels and run as described in section 2.6.2 before immunoblotting as described in section 2.6.3.

To assess if fraction sizes correlate with those estimated by the protein standards, 30μl of every fourth fraction was combined with 10μl of 4X LDS sample buffer and not heated, to retain crosslinks. 30μl of each sample was loaded onto either 4-12% Bis-Tris NuPAGE gels or 3-8% Tris-acetate gels and run as described in section 2.6.2. SYPRO ruby staining was performed as described in section 2.6.4.

2.8.5 LC-MS/MS and analysis of spectra

CBQCA quantitation assay was performed on the peptides as described in section 2.17. From these results, the injection volume for the strongest fraction was calculated to ensure a maximum of 2μg of peptide would be injected. The injection volumes were kept constant for all fractions to ensure elution profiles could be calculated accurately. Using a Thermo Fisher Scientific Ultimate 3000 nanoHPLC system, up to 2μg of peptides in 5% (v/v) formic acid (~10μl) was injected onto an Acclaim PepMap C18 nano-trap column Thermo Fisher Scientific.

After being washed with 2% (v/v) acetonitrile/0.1% (v/v) formic acid, peptides were resolved on a 50cm EASY-Spray column over a gradient from 2% acetonitrile to 35% acetonitrile over 140 minutes with a flow rate of 200nl min^{-1}. The peptides were ionized by nano-electrospray ionization at 2kV. Tandem mass spectrometry analysis was carried out on a Q Exactive mass spectrometer (Thermo Fisher Scientific) using higher energy collisional dissociation (HCD) fragmentation. Precursor peptide and fragment ion measurement were performed in the Orbitrap at 60,000 and 15,000 resolution respectively. Data-dependent acquisition was performed with fragments of the top ten most intense peptide ions, a (TOP10) method described previously [Haas et al., 2006]. The RAW data produced by the mass spectrometer were analysed using the quantitative proteomics software MaxQuant [Cox and Mann, 2008] (version 1.4.1.2). This version of MaxQuant includes an integrated search engine, Andromeda [Cox et al., 2011]. The database supplied to the search engine for peptide identifications was the human UniProt database (May, 2014). The mass tolerance was set at 7ppm for precursor ions, and the MS/MS mass tolerance was set at 20ppm. The enzyme was set either as LysC, or trypsin, with up to two missed cleavages. Deamidation of Asn and Gln, oxidation of Met, pyro-Glu (with N-term Gln) were set as variable modifications. N-ethylmaleimide on Cys was searched as a fixed modification. Identification was set to a false discovery rate of 1%. To achieve reliable identifications, all peptides identified were accepted based on the criteria that the number of hits in the “forward” database was at least 100-fold greater than the number of “reverse” database hits, thus resulting in a false discovery rate of less than 1%. The output from MaxQuant provided peptide level data as well as protein level data, grouped by protein isoforms.

2.8.6 Data analysis of crosslinked samples

All data interpretation for this study was performed primarily using the R language (version 2.15.1). The following data analysis was performed by Dr Mark Larance. To create an elution profile for an individual protein isoform in each of the three replicates, the number of peptides observed in each fraction was summed to generate a count of peptides per fraction (spectral counts). The resulting spectral count profile minima and maxima were normalized within the limits of 0 and 1, respectively. The average of the three elution profiles was based on the normalized spectral peptide counts for each fraction. From the three biological replicates, it was required that a protein be identified in all of the replicates with a minimum of two peptides in each. Proteins labelled as either contaminants or reverse hits were removed from the analysis. These stringent criteria left a subset of 6,425 proteins. For each protein, 144 quantitative measurements were recorded across the three replicates and 48 SEC fractions.

To analyse reproducibility between replicates, the peptide count profile across the 48 fractions for each replicate was correlated between replicates. The normalised spectral count profiles across the 48 fractions for each biological replicate were hierarchically clustered across the 144 measurements, based on Elucidean distance measurement and a “complete” agglomeration method. The output of the clustering was presented in a heatmap using the RColorBrewer library. The biological replicates were collapsed by averaging to obtain a resultant mean peptide count profile describing the behaviour of each protein across the 48 fractions. The standard deviation for the three biological replicates per fraction was also calculated to provide standard error bars on the protein profile graphs.

Subcellular localisation analysis was performed by Dr Yasmeen Ahmad. For this analysis, a published dataset in which proteins were separated by subcellular fractionation was used as a reference library [Larance et al., 2013]. A protein was designated to the subcellular compartment where 80% or more of the total protein was detected. Proteins distributed more generally across compartments were excluded from the analysis. Proteins identified in the native and the crosslinked datasets were then compared to the allocable proteins in the reference dataset.

2.9 Cloning, expression and purification of GST-MRG

2.9.1 Cloning of GST-MRG domain

- pGEX4T1 (GST-empty vector) (GE healthcare)
- The MRG insert was generated by Genscript by gene synthesis, encoding amino acids 191-362 of the MORF4L1 protein

(full sequence: H_2N-

NRVEVKVKIPEELKPWLVDDWDLITRQKQLFYLPAKKNVDSILEDYANYKKSRGNTDNKE

YAVNEVVAGIKEYFNVMLGTQLLYKFERPQYAEILADHPDAPMSQVYGAPHLLRLFVRIG

AMLAYTPLDEKSLALLLNYLHDFLKYLAKNSATLFSASDYEVAPPEYHRKAV-COOH)

LR reaction: 150ng MRG construct and 150ng pDEST15 plasmid (GST-vector) (Life Technologies) were combined and made up to 8μl with TE buffer. 2μl of LR Clonase II Enzyme Mix (Life Technologies, 56485) was added and incubated at room temperature for 1 hour. 1μl of proteinase K (Life Technologies, Y02759) was then added and the reaction was incubated at 37^{o}C for 10 minutes.

Transformation: Two aliquots of One Shot BL21 (DE3) chemically competent cells (Invitrogen, C6000-03) were thawed on ice for 5 minutes. 1μl of pGEX4T1 vector and 5μl of the pDEST15MRG LR reaction were added to separate bacterial vials and mixed gently by tapping. The bacteria were incubated on ice for 30 seconds before subjecting them to heat shock at 42^{o}C in a water bath for 30 seconds. The reactions were returned to ice. 250μl of SOC medium (Life Technologies) at room temperature was added to the bacteria and mixed gently by tapping. Bacteria were then plated out onto LB plates containing ampicillin at 100μg/ml and incubated at 37^{o}C overnight

2.9.2 Materials and solutions for protein expression

Induction agent:

1M IPTG (VWR, 437145X) in Milli-Q H_2O

Culture medium:

LB media supplemented with 100μg/ml ampicillin

Pellet re-suspension buffer:

PBS containing 1 x Complete EDTA-free protease inhibitor tablet per 50ml.

Lysozyme:

40mg/ml lysozyme (Sigma, L6876)

50% glycerol

in PBS

Triton X-100:

10% Triton X-100

in PBS

Glutathione Sepharose 4B Beads:

Supplied as 80% slurry glutathione sepharose in 20% ethanol (Amersham Biosciences)

Elution buffer (50ml):

50mM Tris-HCl, pH8.0

154mg reduced glutathione (Sigma, G4251)

CNBr coupling buffer:

0.1M $NaHCO_3$

0.5M NaCl

pH adjusted to 8.3

CNBr rehydration buffer:

10mM HCl

RIPA buffer:

0.1% SDS

0.5% sodium deoxycholate

1% NP-40

50mM Tris-HCl pH 7.4

150mM NaCl

TE buffer:

10mM Tris-HCl (pH8.0)

0.1mM EDTA

2.9.3 GST-fusion protein purification, cell lysis and binding to glutathione

Single colonies of BL21 bacteria expressing GST-empty vector and GST-MRG were inoculated separately into 5ml of culture medium and left to grow overnight at 37^{o}C with shaking. 5ml of each culture was then used to inoculate 500ml of culture medium in separate 2.5L conical flasks. The cultures were grown for ~5 hours at 37^{o}C with shaking until the optical density at 600nm reached 0.6-0.8. Protein production was then induced by the addition of IPTG to a final concentration of 0.1mM and incubated at 37^{o}C for 3-6 hours. Bacterial cells containing the recombinant proteins were harvested by centrifugation at 5,250g in a Beckman Coulter SX4750 centrifuge at 4^{o}C

for 30 minutes with slow deceleration. The supernatants were discarded and the pellets were frozen at -20^{o}C overnight.

The pellets were resuspended in 50µl ice-cold PBS containing protease inhibitors for every 1ml of culture volume (25ml each). Lysozyme was added to a 1mg/ml final concentration and then incubated at 4^{o}C with rotation for 30 minutes. Triton X-100 was then added to a 1% final concentration and the cells were sonicated on ice for a 30 second burst at 10% power (Branson Digital Sonifier 250). The cells were then incubated at 4^{o}C with rotation for 30 minutes. The cells were sonicated again for 2 x 30 seconds at 10% power to ensure cell lysis was complete. The lysates were then subjected to centrifugation at 12,000g (Beckman Coulter Avanti J25 centrifuge, rotor JA25.5) for 30 minutes at 4^{o}C to pellet cellular debris.

For each 500ml of culture, 2ml of glutathione beads (~3ml slurry) were washed in 50ml of PBS to pre-equilibrate the beads. Beads were pelleted by centrifugation at 2,000g for 2 minutes at room temperature with slow deceleration (Beckman Coulter SX4750 centrifuge). The protein lysates were added to the prepared beads and incubated overnight at 4^{o}C on a rotating wheel. Beads were pelleted again at 2,000g for 2 minutes with slow deceleration before being washed twice with 50ml of 1X PBS, all at 4^{o}C. The beads were transferred into a plastic filter column then washed in 30ml of 1X PBS at room temperature. The GST-fusion proteins were then eluted in 10ml elution buffer and collected in 15ml Falcon tubes.

2.9.4 Coupling protein to CNBr-activated Sepharose 4B

The GST-MRG and GST-empty vector proteins were buffer exchanged into CNBr buffer using 3 x 15 minute centrifugations at 4,000g (Beckman Coulter SX4750 centrifuge) through 10kDa molecular weight cut off Amicon Ultra tubes (Millipore). A Bradford

protein quantification assay was performed on the purified proteins and aliquots of each protein were made up to 50μl of both 0.01mg/ml and 0.1mg/ml in 1X LDS sample buffer with 25mM TCEP. 10μl of each were loaded per lane for SDS-PAGE separation as described in section 2.6.2. The gel was then stained with InstantBlue (Expedeon) as described in section 2.6.5, to assess protein purity. Sufficient protein was purified to make 1.2ml of 1 mg/ml of protein coupled beads. Therefore 700μg of CNBr-activated sepharose 4B beads (GE Lifesciences, 17-0430-01) were rehydrated in 50ml 10mM HCl for 30 minutes. The beads were pelleted at 2,000g for 2 minutes at room temperature with slow deceleration (Beckman Coulter SX4750 centrifuge). The CNBr-sepharose was washed three more times in 50 ml 10mM HCl using a steri-cup (Millipore) under vacuum. The CNBr-sepharose beads were washed in 50ml CNBr coupling buffer in the steri-cup under vacuum. The CNBr sepharose beads were then resuspended in 10ml CNBr buffer and divided between two 15ml falcon tubes, pelleted and the buffer removed. 1.2mg of protein in solution (GST-empty vector and GST-MRG), were added to the beads separately and incubated with rotation overnight at room temperature. The beads were then pelleted at 2,000g for 2 minutes at room temperature with slow deceleration (Beckman Coulter SX4750 centrifuge). The CNBr-sepharose was subsequently blocked with 100mM Tris-HCl pH 7.4 for 8 hours at room temperature with rotation. The beads were then pelleted and washed 3 times for 5 minutes in 20ml RIPA buffer. The beads were finally stored as 50% slurries in 1X TBS with 0.02% NaN_3 at 4^oC.

2.10 Immunoprecipitation

2.10.1 Solutions

Immunoprecipitation buffer:

50mM Tris-HCl pH 7.4

150mM NaCl

5mM NEM

1% NP-40

10% glycerol

10mM EDTA

1x EDTA free-Complete Protease inhibitor tablet per 50ml buffer

1x PhosStop tablet per 10ml buffer.

2.10.2 GFP-immunoprecipitation protocol

Four 15cm dishes of U2OS cells expressing LAP1-MRFAP1 were grown to ~50% confluence. One dish was treated with 10μl DMSO, one was treated with 10μg/ml doxycycline (Sigma, D9891) for 18 hours, one dish was treated with 10μg/ml doxycycline and 1μM MLN4924 for 18 hours and the final dish was treated with 10μg/ml doxycycline for 18 hours and 10μM MG132 for 2 hours prior to harvest. Cells were washed twice with ice-cold PBS. 1ml of ice-cold IP buffer was added to each plate and cells were harvested by scraping. The lysates were sonicated, 3 x 10 seconds at 10% power (Branson Digital Sonifier), on ice, before centrifugation at 17,000g for 10 minutes at 4^{o}C (Heraeus Pico17 centrifuge). BCA protein quantitation assay was performed on the supernatant as described in section 2.16. 40μl of GFP-trap agarose bead slurry (ChromoTek) per IP was washed in 1ml IP buffer. Lysate containing 1.5mg of total protein was added to the pre-washed beads and mixed by rotation for 2 hours at 4^{o}C. Beads were pelleted by centrifugation at 2,000g for 2 minutes at 4^{o}C and the

flow through was retained. The beads were washed three times in 1ml IP buffer. The beads were resuspended in 200μl ice-cold PBS per IP and transferred to a spin column. Beads were dried by centrifugation at 500g at 4°C. The spin columns were transferred to clean Eppendorf tubes and the immunoprecipitated proteins were eluted by the addition of 50μl 1X LDS sample buffer supplemented with 25μM TCEP pre-heated to 65°C. The samples were heated to 65°C for 5 minutes then centrifuged at 500g for 5 minutes at room temperature. BCA protein quantitation was performed on the flow through, and both the start and the flow through were made up to a final concentration of 1mg/ml of total protein with 1X LDS sample buffer and 25μM TCEP for immunoblotting.

2.10.3 Endogenous immunoprecipitation

One 15cm dish of cells per IP was grown to 80% confluence. For VprBP and MORF4L1 IPs, either 1μM MLN4924, or 10μl DMSO control, was added 16-20 hours prior to harvest. Cells were washed twice in ice-cold PBS before scraping cells in 1ml IP buffer per dish. Lysates were sonicated for 3 x 10 seconds, 10% power on ice (Branson Digital Sonifier 250). Lysates were then centrifuged at 17,000g for 10 minutes at 4°C (Heraeus Pico17 centrifuge, Thermo Fisher Scientific). BCA quantitation assay was performed on the supernatant (section 2.16). Protein G Dynabeads (Invitrogen, 10003D) were vortexed vigorously and 50μl of slurry per IP was washed in 1ml IP buffer. The beads were pelleted on a Dynal magnet (Life Technologies) and then resuspended in 100μl of IP buffer per IP and divided into 1.5ml Eppendorf tubes, one for each IP. Lysate containing 1.5mg total protein was mixed with either 10μg of the rabbit IgG control antibody (Jackson Immuno Research Laboratories, 011-000-003), or 10μg MORF4L1

antibody (Table 2.1). Lysate containing 1mg total protein was mixed with either 3μg of the rabbit IgG control antibody, or 3μg VprBP antibody (Table 2.1).

The antibody mixtures were added to the Dynabeads and mixed with rotation for 2 hours at 4°C. The beads were then pelleted on a magnet and the flow through was retained. Beads were washed three times with 1ml IP buffer at 4°C. Beads were then washed in 1ml ice-cold PBS. Protein was eluted from the beads by incubating in 1xLDS sample buffer containing 25mM TCEP at 65°C for 5 minutes. Beads were separated from the eluate by pelleting on a magnet and the eluate was then transferred to a clean 1.5ml Eppendorf tube. Protein content in the start and flow through material was quantified by BCA assay and an aliquot was made up to 1mg/ml total protein with 1xLDS sample buffer and 25mM TCEP for immunoblotting.

2.10.4 GST-MRG pull down (non-SILAC)

Three 15cm dishes of LAP1-7KR MRFAP1 mutant U2OS cells were grown to 80% confluence. The cells were washed twice with 20ml of ice-cold PBS. All three plates were scraped in a total volume of 1.5ml IP buffer. The lysate was sonicated 3 x 10 seconds at 10% power (Branson Digital Sonifier 250), on ice, then centrifuged at 17,000g at 4°C (Heraeus Pico17 centrifuge, Thermo Fisher Scientific). BCA assay was performed on the supernatant (as described in section 2.16). Forty micrograms of bead bound GST-MRG protein per IP was washed in 1ml IP buffer, pelleted at 2,000g for 2 minutes (Beckman Coulter SX4750 centrifuge). The beads were resuspended in 500μl IP buffer and divided between five clean Eppendorf tubes. 4mg, 2mg, 1mg, 0.5mg of lysate protein, or just IP buffer alone, was added to the prepared beads and mixed with rotation overnight at 4°C. The beads were washed and protein eluted as described above in section 2.7.2. BCA assay was performed on the flow through.

Aliquots of the starting material and flow through were made up to 1mg/ml with 1X LDS sample buffer and 25mM TCEP for SDS-PAGE and immunoblotting as described in section 2.6.

2.10.5 GST-MRG SILAC immunoprecipitation

LAP1-7KR MRFAP1 mutant U2OS cells were grown in medium containing R0K0, R6K4 and R10K8 labelled amino acids for three passages (1:6 split) to ensure maximum incorporation of labelled amino acids. Two 15cm dishes of each were grown to 50% confluence and two plates of R10K8 labelled cells and one of the R6K4 plates were induced with doxycycline (10μg/ml) for 96 hours prior to harvest. For harvesting, cells were scraped in 500μl IP buffer per plate, on ice. The lysate was sonicated on ice for 3 x 10 second bursts at 10% power (Branson Digital Sonifier 250) then centrifuged at 17,000g for 10 minutes at 4^{o}C (Heraeus Pico17 centrifuge, Thermo Fisher Scientific). BCA protein quantitation assay was performed on the supernatant and 4mg of protein per treatment group was used for the pull down. 40μg of bead-bound protein per condition were washed in IP buffer. Non-induced R0K0 cell lysates were incubated with GST-MRG beads as a control and induced R10K8 cell lysates were also incubated with GST-MRG beads as the treatment group. The part induced/part non-induced R6K4 lysate was mixed with GST-empty vector beads as a control for non-specific, GST/bead interacting proteins. The lysates were mixed with the beads overnight with rotation at 4^{o}C. Beads were pelleted by centrifugation at 2,000g with slow deceleration (Beckman Coulter SX4750 centrifuge), at 4^{o}C and the flow through was retained. The beads were then combined and washed together in 1ml of IP buffer at 4^{o}C. Two further washes were performed using the same buffer. The beads were then resuspended in 500μl of ice-cold PBS and transferred to a spin column and centrifuged at 500g for 1 minute at

4°C (Heraeus Pico17 centrifuge, Thermo Fisher Scientific). Protein on the dried beads was eluted by the addition of 50µl of 1XLDS sample buffer with 25mM TCEP (pre-heated to 65°C). The eluate was collected by centrifugation at 500g for 1 minute into a clean Eppendorf tube. 25µl of elution was loaded onto two side-by-side lanes of a 4-12% 10 well 1mM NuPAGE Bis-Tris gel. The gel was run with MES buffer at 200V for 43 minutes. The gel was stained with SYPRO Ruby protein stain as per section 2.6.4.

2.11 In-gel digestion

2.11.1 Solutions

- 0.5M NH_4HCO_3 (50ml)
- 0.1M NH_4HCO_3 (50ml)
- 50:50 – 50% CH_3CN, 0.25M NH_4HCO_3 (50ml)

All stock solutions were made fresh each day and were made using HPLC grade H_2O (VWR).

- 5% Formic Acid
- 50mM Acetic acid

2.11.2 Protocol

The slightly wet, SYRO Ruby stained gel was placed on a clean flat surface (lid of a 15cm tissue culture plate) and gel bands were viewed under UV light. Gel bands were cut into 13 slices from low to high molecular weight. Each slice was cut into 1mm cubes and transferred to a Low-bind Eppendorf tube. The gel cubes were washed once in 1ml of 50% CH_3CN/0.25M NH_4HCO_3 with slow vortexing for 30 minutes at room temperature. Gel bands were dehydrated by the removal of all liquid and the addition of 1ml of 100% CH_3CN with shaking for 10 minutes at room temperature. All liquid was

carefully removed. Trypsin (MS-grade, Roche) was suspended in 50mM acetic acid to a final concentration of 1μg/ml. Trypsin was diluted to 12.5ng/μl in 100mM NH_4HCO_3 and 30-40 μl was added to each tube. Tubes were centrifuged at 17,000g for 1 minute at room temperature (Heraeus Pico17 centrifuge, Thermo Fisher Scientific), then incubated overnight at 37°C. 100μl 5% formic acid was added to each tube, vortexed and incubated for 1 hour at 37°C. 100 μl 100% CH_3CN was added to each tube, vortexed and incubated for 1 hour at 37°C. 500 μl 100% CH_3CN was added to each tube, vortexed and incubated for 10 minutes at 37°C. The supernatant was moved to a new low-bind Eppendorf tube and dried at 30°C overnight using a Savant SPD 2010 SpeedVac concentrator (Thermo Fisher Scientific). The dried peptides were resuspended in 20μl 5% formic acid, vortexed and incubated at room temperature for 30 minutes. CBQCA peptide quantitation was performed as described in section 2.17. 1μg of peptide from each of the 13 vials was used for MS analysis.

2.11.3 MS-analysis of SILAC pull down samples

Using a Thermo Fisher Scientific Ultimate 3000 nanoHPLC system, a maximum of 1μg of peptides in 5% (v/v) formic acid was injected for each gel slice onto an Acclaim PepMap C18 nano-trap column (Thermo Fisher Scientific). After being washed with 2% (v/v) acetonitrile, 0.1% (v/v) formic acid, peptides were resolved on a 150mm × 75μm Acclaim PepMap C18 reverse-phase analytical column over a gradient from 2% acetonitrile to 80% acetonitrile over 100 minutes with a flow rate of 300nl min^{-1}. Thereafter, ionisation and MS analysis was performed as described in section 2.7.7 (SEC non-crosslinked).

2.12 Cell immunofluorescence

2.12.1 Solutions and materials

- 16% methanol free paraformaldehyde (PFA) (Pierce, 28908)
- Vectashield (Vector Laboratories, H-1400)

Fixation Buffer:

3% PFA in PBS, made fresh for each use

Quench Buffer:

50mM Glycine in PBS

Blocking Buffer:

5% BSA

in TBS and 0.1% Tween-20

Permeabilisation buffer:

1% Triton X-100 in PBS

2.12.2 Immunofluorescence protocol

Cells were grown on 13mm coverslips in 24 well plates. For immunofluorescence analysis, medium was aspirated and cells were washed in 1ml PBS per well. The cells were fixed in 1ml fixation buffer per well for 20 minutes at room temperature on a rocker, then washed in 1ml PBS per well. To block excess PFA, the cells were incubated in 1ml quench buffer per well for 5 minutes at room temperature on a rocker. Cells were permeabilised with 1ml permeabilisation buffer per well, for 10 minutes at room temperature on a rocker. The coverslips were then washed in PBS (1ml/well). Blocking buffer (1ml/well) was then added to the coverslips and incubated at room temperature for 30 minutes on a rocker. Blocking buffer was removed and then primary antibodies (see Table 2.1) were added to the coverslips at the appropriate dilutions, diluted in blocking buffer, and incubated for 1 hour at room temperature.

Coverslips were washed in PBS (1ml/well) three times, for 5 minutes each, on a rocker at room temperature. Alexa-Fluor-labelled secondary antibodies (Table 2.1) were prepared in blocking buffer at 1:400 dilution. 500µl of diluted antibodies were added to each coverslip. DAPI (1:1000) was added to the Alexa-Fluor antibodies and incubated for 30 minutes at room temperature on a rocker, covered from light. Coverslips were washed in PBS (1ml/well) 3 times for 5 minutes, on a rocker. Control slides were incubated with DAPI and secondary antibodies only. Slides were prepared by placing a drop of Vectashield mounting medium for each coverslip. Coverslips were then placed, cell side down, on the mounting medium and left in the dark at room temperature for 24 hours. The coverslips were then sealed with nail varnish and dried at room temperature, protected from light.

2.13 Tissue immunofluorescence

2.13.1 Solutions

Blocking buffer:

1X PBS

5% goat serum (Sigma)

0.3% Triton X-100

Antigen Unmasking buffer:

Tris base 1.2g

EDTA 0.37g

in 1L H_2O, pH adjusted to 9 with HCl

Permeabilisation buffer:

1% triton X-100 in PBS

Sodium Borohydrate:

1% (w/v) in PBS

Quench buffer:

50mM Glycine in PBS

2.13.2 Immunofluorescence protocol

To remove paraffin wax, sections were incubated in Histo-clear (National Diagnostics) for 2 x 5 minutes. Sections were rehydrated in 100% ethanol for 2 x 5 minutes then 70% ethanol for 5 minutes. The sections were then rinsed in running tap water for ~1 minute. Antigen retrieval was performed by heating sections immersed in antigen unmasking buffer for 5 minutes in the microwave at full power. Sections were then permeabilised in 100μl permeabilisation buffer for 15 minutes then rinsed with 3 x 1ml PBS. Sections were then incubated in sodium borohydrate solution for 4 x 15 minutes with the solution made freshly immediately before each incubation. Sections were rinsed 3 x in PBS then quenched in quench buffer for 15 minutes. Sections were blocked in 100μl blocking buffer for 60 minutes then incubated with primary antibodies (Table 2.1) and DAPI 0.5μg/ml (Sigma), diluted in blocking buffer. Primary antibodies were washed with 3 x 1ml PBS before incubation with fluorochrome-conjugated secondary antibodies diluted in blocking buffer at 1:400 for 1 hour at room temperature in the dark. Control sections were incubated with DAPI and secondary antibodies only. Sections were washed in 3 x 1 ml PBS, dried then covered with Vectashield mounting medium under coverslips.

2.14 Microscopy.

Microscopy was performed on a Leica SP2 confocal scanning fluorescence microscope. The pinhole was set to 1 Airy unit. Lasers used were the 405nm blue diode, 488nm argon ion laser and 594nm orange diode laser. A 63x objective lens plan Apo was used with 1.3 numerical aperture. Images were analysed using Leica LCS Software.

2.15 Chloroform methanol precipitation

This method is useful for concentrating dilute protein solutions in any buffer. SDS was added to the protein solution to a final concentration of 2% and heated to 65^{o}C for 10 minutes to denature proteins. This assists solubilisation of the pellet in the last step. 400µl of 100% methanol was added to 100µl of protein containing solution (including SDS) and vortexed for 5 seconds. 100µl of 100% chloroform was added and vortexed again. 300µl MilliQ water was added and vortexed for 1 minute. The sample was centrifuged at 9,000g for 5 minutes at room temperature (Heraeus Pico17 centrifuge, Thermo Fisher Scientific). The top 80% of the upper phase was discarded. 300µl of 100% methanol was added to the tube and vortexed briefly. The sample was centrifuged at 17,000g for 5 minutes at room temperature. The entire supernatant was removed and the pellet air dried at room temperature. The pellet was then resuspended in the desired volume of 1X LDS sample buffer containing 25mM TCEP to adjust the final concentration to 1mg/ml.

2.16 Bicinchoninic acid (BCA) quantitation assay

2.16.1 Solutions and materials

- BCA protein assay kit (Pierce, PI-23227)
- 1X PBS
- BSA protein standards prepared in PBS to final concentrations of: 0.025, 0.075, 0.2, 0.4, 0.6, 0.8, 1 and 2mg/ml

2.16.2 Protocol

25µl of PBS (blank) and each protein standard were loaded in triplicate into a flat bottomed, clear, 96-well plate (Thermo Fisher Scientific). Proteins samples were

diluted 10 or 20 fold in 1X PBS and 25µl loaded into wells in duplicate or triplicate. 200µl of BCA working reagent was added to each well, covered and then incubated at 37^{o}C for 30 minutes. The absorbance was then measured at 562nm on a SpectraMax M2E microplate reader (Molecular devices) and SoftMax Pro software (Molecular Devices) was used to analyse the results.

2.17 CBQCA peptide quantitation assay

2.17.1 Solutions and materials

- 0.1M H_3BO_3-NaOH pH 9.3 (Borate buffer)
- CBQCA protein quantitation kit (Invitrogen, C-6667)
- Black bottom 96-well plate for fluorescence with lid (Thermo Fisher Scientific)
- BSA peptide standards, trypsin digested, ultrafiltered with 10kDa cut-off, C18 Sep-pak purified, in borate buffer, in Eppendorf 1.5ml low protein binding tubes at concentrations of 655360ng/ml, 163840ng/ml, 40960ng/ml, 10240ng/ml, 2560ng/ml and 640ng/ml.

2.17.2 Protocol

The assay was performed as per manufacturer's instructions. Briefly, peptides were diluted 25-fold in borate buffer in the 96-well plates. Borate buffer blank wells and the standards were added to the plate in triplicate. For each well, 95µl of borate buffer was mixed with 5µl of KCN solution and 100µl of this mixture was added to each well. 25µl of a 2mM CBQCA reagent stock was added to each well. The plate was covered with a lid and incubated on a rocker at room temperature for 1 hour. The fluorescence was then read by excitation at 465nm and emission at 550nm with 495nm cut off filter

using a SpectraMax M2E microplate reader (Molecular devices) and SoftMax Pro software (Molecular Devices) was used to analyse the results.

2.18 Coomassie Plus (Bradford) quantitation assay

2.18.1 Solutions and materials

- BSA standards as per section 2.7.1
- Coomassie Plus Reagent (Thermo Fisher Scientific, 23236)

2.18.2 Protocol

25µl of PBS blanks and BSA protein standards were loaded in triplicate onto a flat bottomed, clear, 96 well plate. 200µl of Coomassie Plus reagent was added to each well and mixed gently. The plate was incubated at room temperature for 10 minutes then the absorbance was measured at 595nm using a SpectraMax M2E microplate reader (Molecular devices). SoftMax Pro software (Molecular Devices) was used to analyse the results.

2.19 EZQ protein quantitation assay

2.19.1 Solutions and materials

- Protein standards as per section 2.7.1
- EZQ Protein quantitation kit (Invitrogen, R33200)

Rinse Buffer:

10% methanol

7% acetic acid

in MilliQ H_2O

2.19.2 Protocol

A sheet of assay paper was fitted to the microplate cassette supplied as part of the assay kit. 1µl of 1X PBS, protein standards and sample were applied to the paper in triplicate. In the case of SEC fractions, the samples were analysed as single replicates only. When required, samples were diluted in the same buffer they were dissolved in. The protein samples were dried completely with the aid of a hair dryer. The paper was fixed and washed with 40ml 100% methanol in a plastic tray with agitation for 5 minutes. The paper was then dried using a hair drier. The paper was stained with 40ml EZQ staining reagent and incubated at room temperature with agitation for 30 minutes, covered from light. After staining, the assay paper was rinsed for 1-2 minutes in rinse buffer for a total of three rinses. The paper was dried using a hair dryer then analysed in a SpectraMax M2E microplate reader (Molecular devices) using excitation/emission settings of ~480/590nm. SoftMax Pro software (Molecular Devices) was used to analyse the results.

Chapter 3: Native protein complex analysis by size-fractionation-based quantitative proteomics

3.1 Introduction

As discussed in Chapter 1, the wide variety of possible protein-protein interactions enables the large diversity of cellular protein functions. While a number of protein complexes have been widely studied, to date very little is known about the composition of many of the protein complexes formed in human cells. The CORUM database forms the largest dataset of protein complexes compiled from available literature, and contains information relating to 1,970 protein complexes identified in human cells from only ~16% of the known protein-coding genes [Ruepp et al., 2010]. Therefore, a huge deficit still exists in our knowledge of the putative complexes formed by the remaining proteins. More detailed knowledge of protein complexes and the cellular pathways they regulate would be invaluable for the life sciences and medical communities.

To identify protein-protein interactions, affinity purification, typically of tagged protein baits, combined with subsequent mass spectrometric analysis of the isolated proteins has been widely used. Although this approach gives vast amounts of detail regarding the interacting proteins, a significant limitation to these techniques is that they do not distinguish the distinct functional complexes the interacting proteins form. It is also impractical in terms of time and cost to apply affinity purification methods on a

system-wide scale and the use of tagged proteins limits applications, for example with clinical samples. To overcome these limitations, some studies have used blue native polyacrylamide gel electrophoresis (BN-PAGE) combined with mass-spectrometry to study native protein complexes within either purified subcellular fractions [Heide et al., 2012], or whole cell lysates [Camacho-Carvajal et al., 2004]. Limitations in this technique, however, include the low resolution for complexes <100kDa, both due to the high abundance of complexes in this size range and the limited separation distance resulting from the acrylamide gradient [Eubel et al., 2005]. In addition, the total protein loading capacity (400μg per 0.16 x 1cm sample well [Wittig et al., 2006]) and reproducibility due to manual implementation of the BN-PAGE technique limit its applicability to large scale experiments. More recently, ion exchange chromatography (IEX) combined with mass spectrometry has also been utilised to study soluble native protein complexes in human cell lines [Havugimana et al., 2012].

Size exclusion chromatography (SEC) is a well-established technique used to separate proteins in solution on the basis of their size (rotational cross-section) [Wheaton and Bauman, 1953]. SEC has been used in previous studies to separate native protein complexes in plant chloroplasts [Olinares et al., 2010] and for the study of large cytosolic complexes in mammalian cells [Kristensen et al., 2012] but to date, this technique has not been extensively applied in combination with mass spectrometry to study all of the native protein complexes in soluble whole cell extracts.

The aim of this chapter was to develop a mass-spectrometry compatible technique that would enable the system-wide analysis of native protein complexes including the analysis of protein isoforms and post-translationally modified peptides.

3.1.1 Hypothesis and aims

The hypotheses for this work were that

- Intact protein complexes can be analysed in a high-throughput, mass spectrometry based approach
- Size exclusion chromatography permits the separation of native protein complexes in a native buffer
- Protein isoforms and post translational modifications interact in different complexes to their unmodified counterparts

To test this hypothesis I aimed:

- To develop a method for the analysis of native protein complexes utilising size exclusion chromatography in combination with tandem mass spectrometry
- To analyse the data at the level of protein isoforms to identify differences in protein elution profiles
- To study whether the elution profiles of post-translationally modified proteins differ from their unmodified counterparts

3.2 Results

3.2.1 Workflow for the identification of native protein complexes

To identify and characterise native protein complexes from human U2OS osteosarcoma cells, lysis was performed using PBS containing protease and phosphatase inhibitors. No detergents were used to prevent the disruption of protein-protein interactions (Figure 3.1). The resulting protein complexes were fractionated by size (rotational cross-section) using non-denaturing size exclusion chromatography

(SEC) and the eluate was collected into 40 sequential fractions (Figure 3.1). Fractions were digested with trypsin, desalted and analysed by LC-MS/MS. Peptides were identified and quantified using the MaxQuant Andromeda software [Cox and Mann, 2008; Cox et al., 2011], using standard parameters, as described in the Methods chapter (section 2.7.7). From the three biological replicates, it was required that a protein be identified in at least two replicates with a minimum of two peptides in each. These stringent criteria left over 71,500 peptides, corresponding to 8,168 proteins, for further analysis.

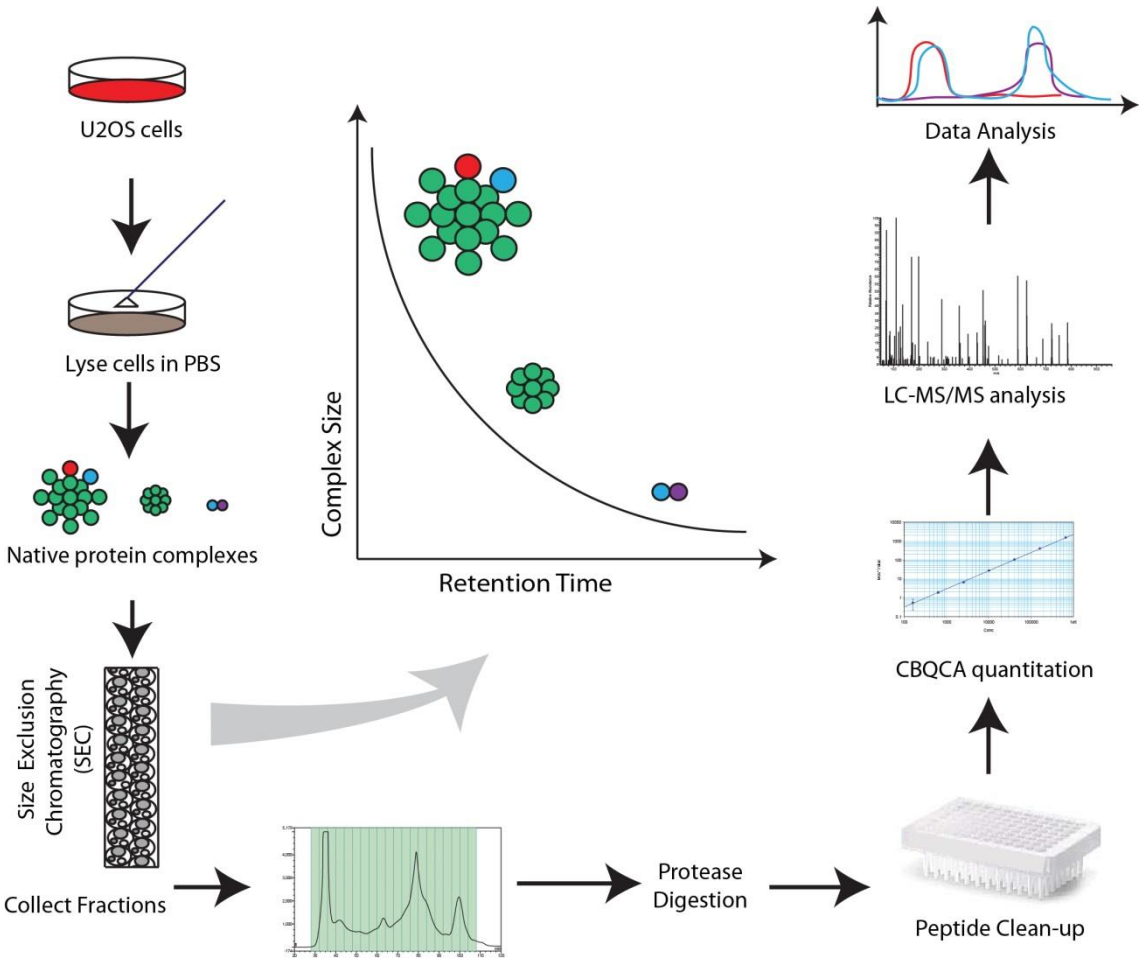


Figure 3.1. Native protein complex analysis using a comprehensive proteomics workflow.

Workflow for SEC-based protein complex separation.

3.2.2 Fraction level analysis of protein properties

Protein standards were analysed using the above method to identify the approximate molecular weights of each fraction. This showed that the method separated protein complexes with molecular weights from ~15kDa to >1MDa (Figure 3.2). For subsequent analysis protein complexes were deemed to be either larger (fractions 7-23), or smaller (fractions 24-40), using a cut-off between 130 and 440kDa (Figure 3.2).

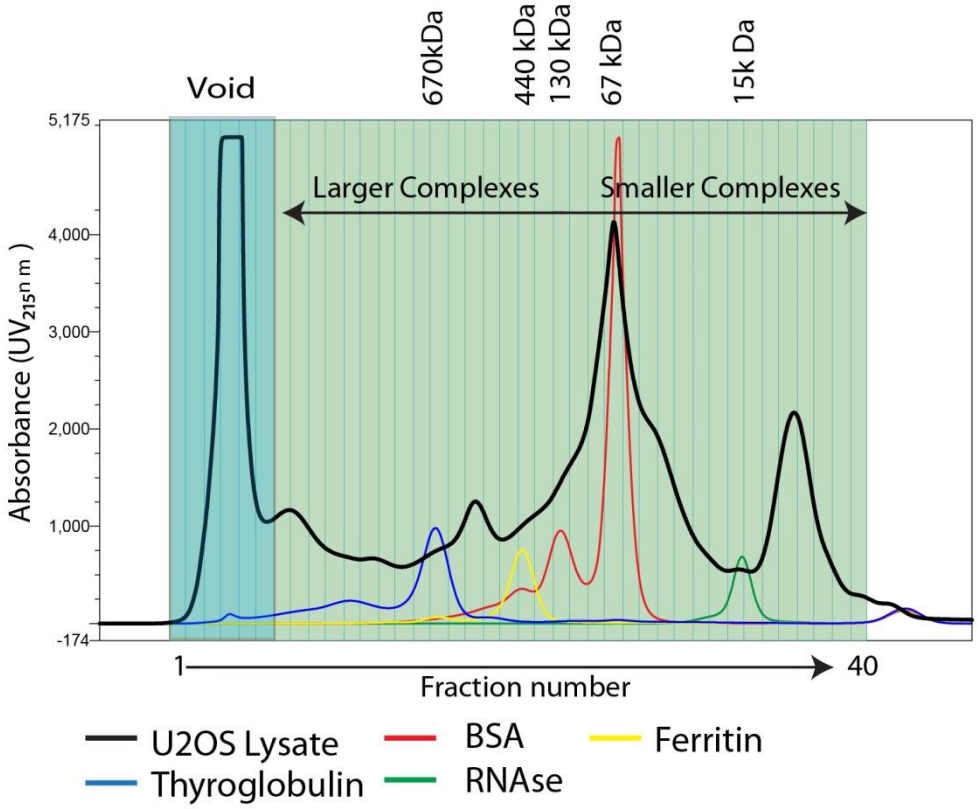


Figure 3.2. Characterisation of protein complex separation.

The UV chromatograph from one of the biological replicates indicates the profile of the U2OS total cell lysate as it eluted from the Superose 6 SEC column across 40 fractions. Protein standards of known molecular weights (thyroglobulin, ferritin, BSA, RNase) were injected onto the same column, and their elution peaks were used to calculate approximate molecular weights for fractions. Proteins that eluted in fractions 1-6 were termed the void. Complexes that eluted in fractions 7-23 were classed as “larger”, and those in fractions 24-40 were classed as “smaller”.

The elution profile for each protein standard was used to calculate a predicted molecular weight for each fraction. Using these predicted molecular weights, a linear regression model was generated (Figure 3.3), which enabled the calculation of an approximate observed molecular weight of complexes/proteins found in the dataset, based on their peak elution fraction.

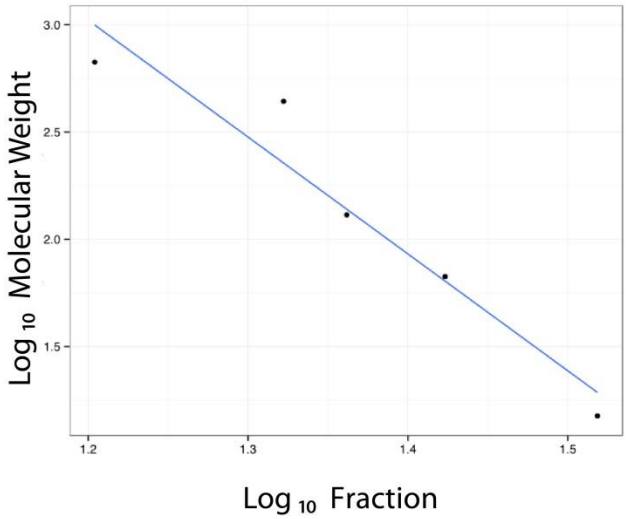


Figure 3.3. Linear regression model to approximate molecular weight of complexes/proteins within each fraction.

The Log_{10} of the SEC fraction number was plotted against the Log 10 molecular weight of the proteins standards to calculate a linear regression (y=9.561209-5.449685x). Analysis performed by Dr Yasmeen Ahmad.

Predicted protein molecular weights were extracted from the Uniprot database [Consortium, 2013] and compared to the calculated molecular weight, based on the linear regression model, for each protein (Figure 3.4). The scatter plot shows that most proteins were generally identified at a higher molecular weight than their Uniprot molecular weight, which supported my hypothesis that the majority of proteins in the human proteome are involved in one or more complexes.

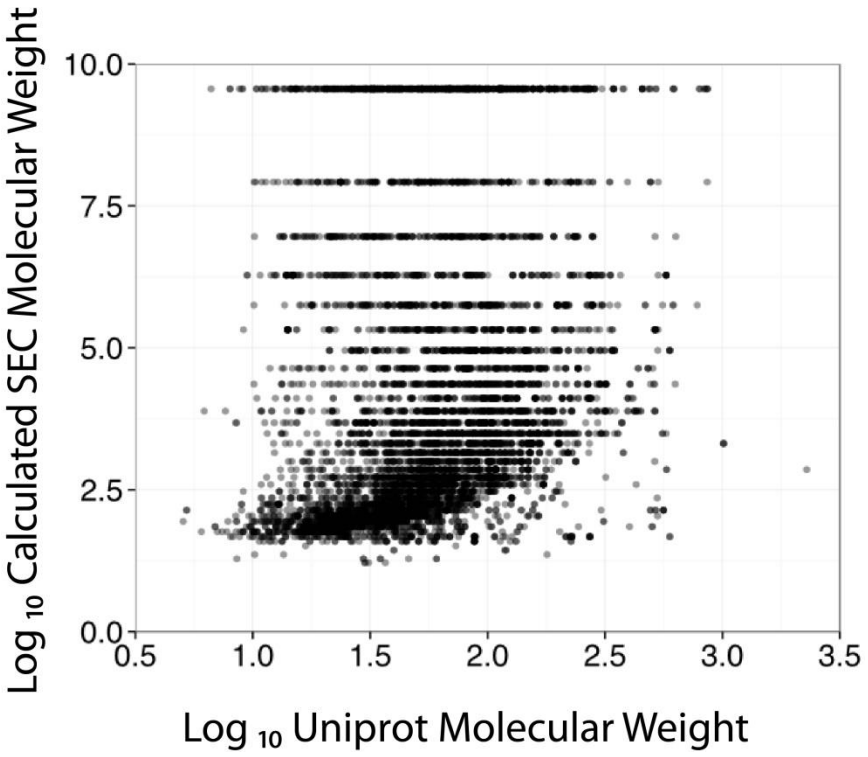


Figure 3.4. Comparison of Uniprot predicted protein molecular weight versus SEC observed protein molecular weight.

The linear regression equation was used to calculate an SEC observed molecular weight for each protein based on the fraction containing its main elution peak. The Log_{10} observed molecular weight (y-axis in kDa) was then plotted against the Log_{10} predicted molecular weight (x-axis in kDa) as a scatter plot. Analysis performed by Dr Yasmeen Ahmad.

To confirm that the SEC method separated intact protein complexes, pairs of consecutive elution fractions were combined (resulting in 20 combined fractions in total), then separated by molecular weight using SDS-PAGE and stained for total protein. This analysis showed that the earliest fractions eluting from the column (larger complexes), contained proteins with very heterogeneous sizes, whereas fractions eluting later (smaller complexes), contained proteins that mostly migrated either at, or close to, their predicted molecular weight (Figure 3.5). This suggests that either small complexes are less heterogeneous, or that more monomeric proteins were isolated in

these fractions. I also observed that the first three fractions (labelled void), contained similar proteins to the total lysate, likely including complexes larger than the separation range of the column. Therefore, these fractions were not included in any further analysis (Figure 3.5).

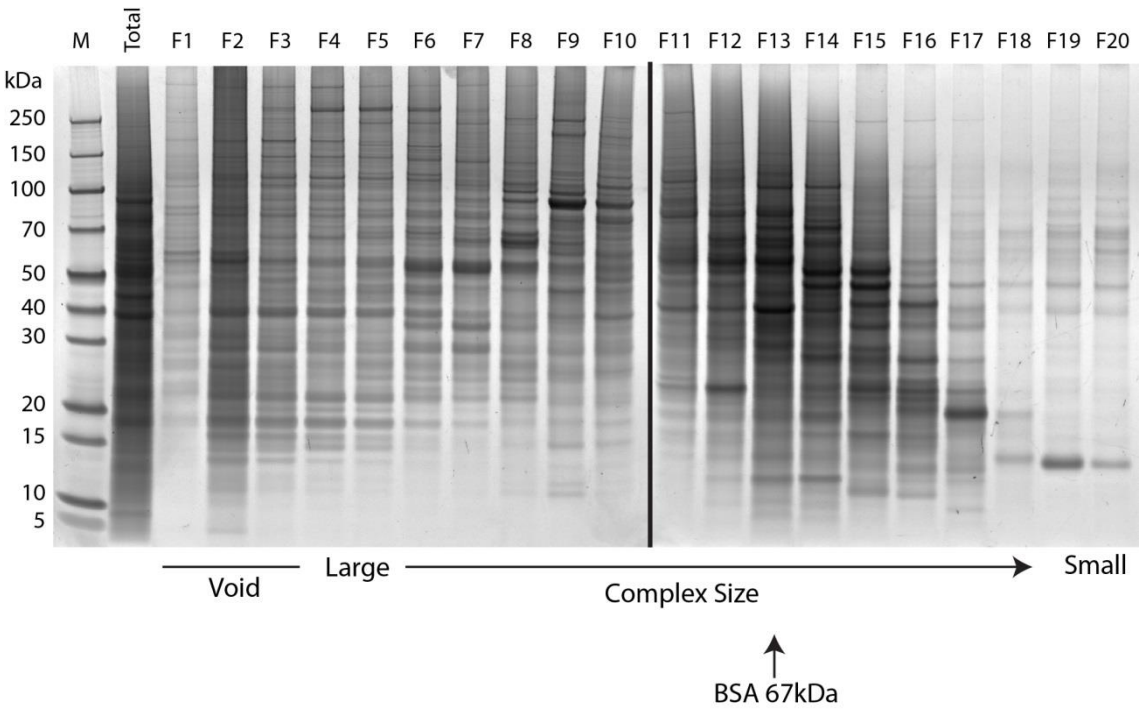


Figure 3.5. Complexes separated by native SEC contain proteins of heterogeneous sizes.

SYPRO ruby total protein stain of proteins in the total cell lysate and consecutively combined fractions. 10μg of protein was loaded in lane 2 (total), and a maximum of 10μg of protein was loaded per lane for elution fractions.

In contrast, proteins isolated from U2OS cells using a highly denaturing buffer, migrate predominantly at their predicted molecular weight when injected into the same SEC column (Figure 3.6).

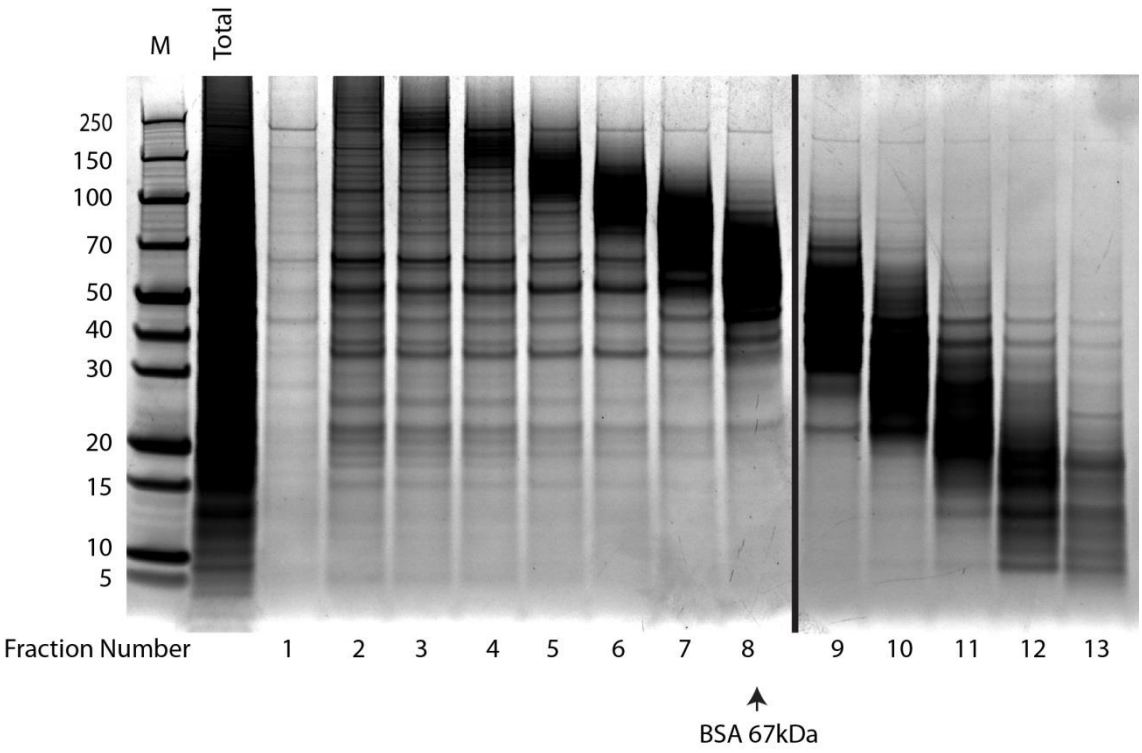
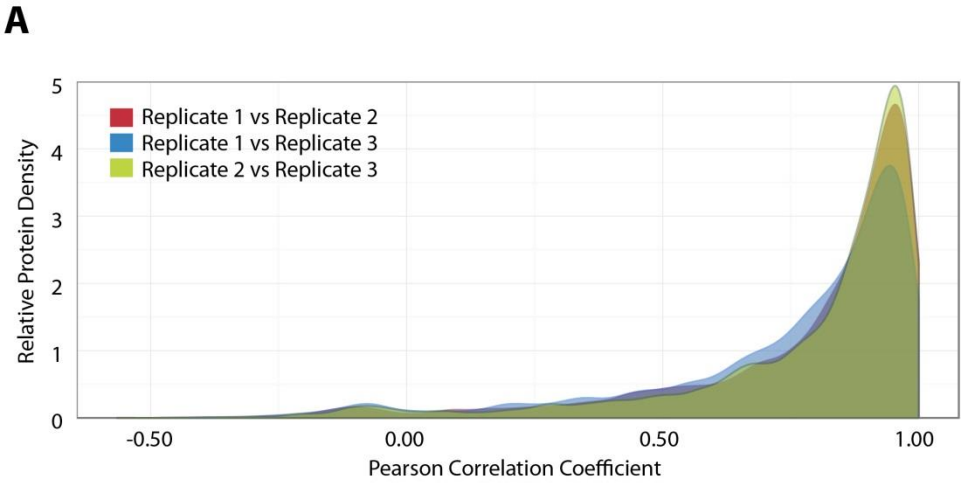


Figure 3.6. Denaturing size exclusion chromatography.

SYPRO ruby total protein stain of proteins in the total cell lysate and consecutively combined fractions. 10μg of protein was loaded in lane 2 (total), and a maximum of 10μg of protein was loaded per lane for elution fractions.

3.2.3 Reproducibility of the SEC elution profiles

To determine the reproducibility of the protein complex separation by SEC, a pairwise comparison was carried out between three biological replicates. In each comparison, a Pearson correlation coefficient was calculated to compare normalised protein profiles across the 40 fractions. Each protein profile was normalised using the maximum peptide count from any fraction within the profile. Each pairwise comparison is shown on a density plot, demonstrating a high correlation (>0.8) for the majority of proteins (Figure 3.7A).


A
Replicate 1 vs Replicate 2
Replicate 1 vs Replicate 3
Replicate 2 vs Replicate 3
Relative Protein Density
Pearson Correlation Coefficient
-0.50
0.00
0.50
1.00

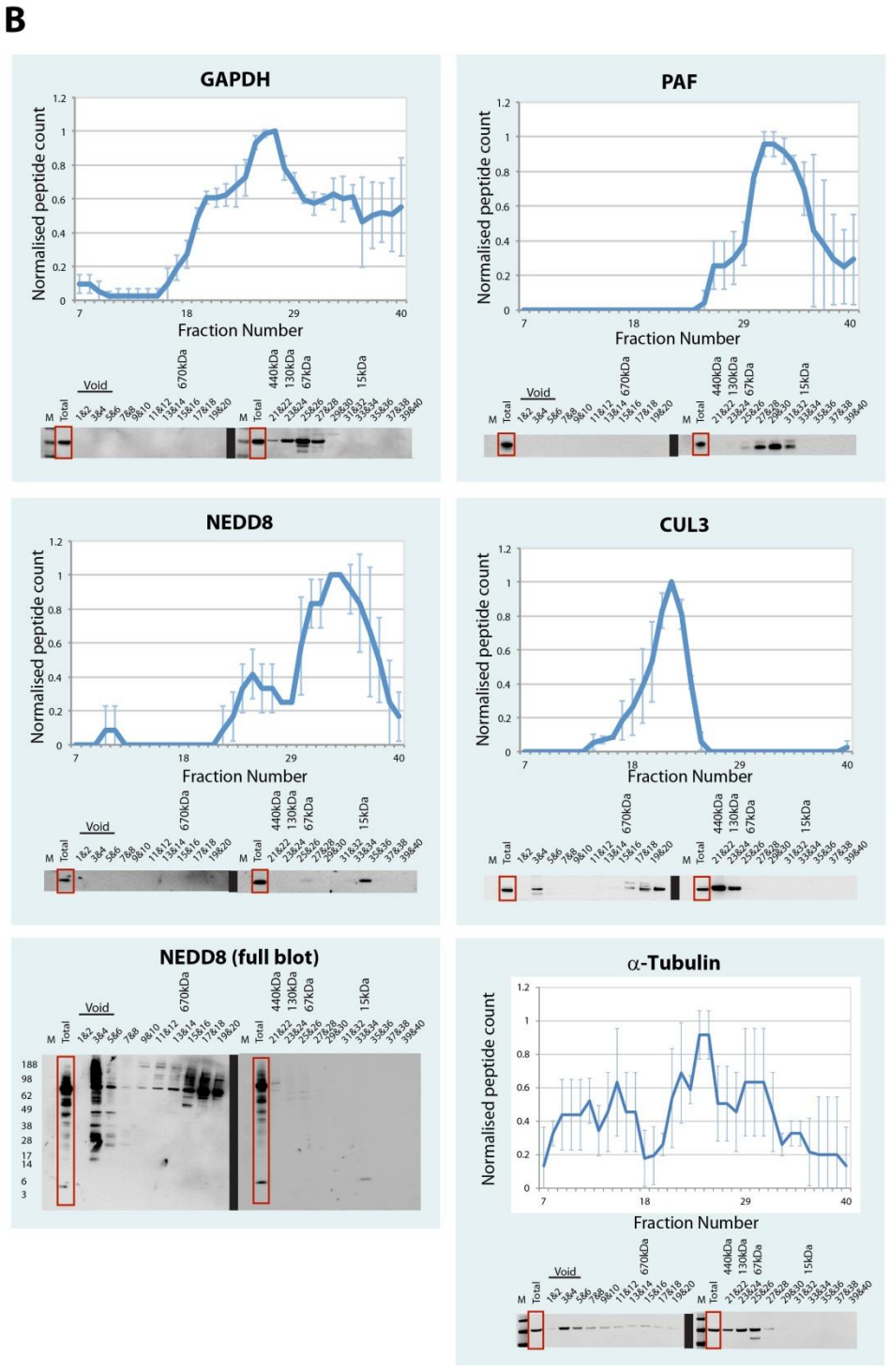

B
GAPDH
PAF
NEDD8
CUL3
NEDD8 (full blot)
α-Tubulin
Normalised peptide count
Fraction Number
Void
670kDa
440kDa
130kDa
67kDa
15kDa

Figure 3.7. Reproducibility of the SEC elution profiles.

A, the Pearson correlation coefficient was calculated for the three biological replicates (replicate 1 *versus* replicate 2, replicate 1 *versus* replicate 3, and replicate 2 *versus* replicate 3). The relative protein density (y-axis) was plotted against the Pearson correlation coefficient for each replicate (x-axis). *B*, the SEC elution profiles for selected proteins are displayed as line graphs with the normalised peptide count (y-axis) plotted against each elution fraction (x-axis). Error bars indicate the standard deviation from the mean across three biological replicates. Below each line graph are immunoblots for each protein of interest. A total of 10μg of protein was loaded in the "total" lane and up to a maximum of 10 μg of protein was loaded per lane for each of the combined fractions. The annotated molecular weights were estimated from the elution profiles of the protein standards injected onto the SEC column.

Many proteins demonstrate one or more 'peaks' in their elution profile, as can be seen in the MS peptide elution profiles for a selected group of proteins (Figure 3.7B). I hypothesised that these peaks correspond to these proteins participating, at least in part, in complexes with other proteins co-eluting within that fraction. Many proteins showed complex elution profiles, as expected. This is because a combination of protein polymers, large and small complexes, as well as monomeric proteins, were detected within the cell lysate.

To characterize further the separation of protein complexes by native SEC, I analysed a variety of proteins using the orthogonal technique of Western blotting. For this, the pairwise pooled SEC fractions, as described above, were separated via SDS-PAGE and immunoblotted to detect five representative proteins with different elution profiles displaying profile peaks in different fractions. Regardless of whether the protein was a small ubiquitin-like modifier (NEDD8), a protein involved in large complexes (CUL3), a protein known to form either multimers (GAPDH), or extended polymers (tubulin), or a small DNA damage repair protein (PAF), there was not only a close match within the

MS replicate data (*i.e.* relatively small error bars), but also a close match between the MS data and the immunoblot profiles (Figure 3.7B).

3.2.4 Analysis of protein and complex coverage

To determine if the dataset was biased, for example against either low, or high, copy number proteins, the data from a previous study of general protein abundance [Nagaraj et al., 2011] was combined with the CORUM database, which contains known protein complexes [Ruepp et al., 2010]. Figure 3.8A shows the normal distribution of protein copy numbers for proteins found in complexes, as suggested by the CORUM database. The same comparison of protein copy numbers was applied to my dataset (Figure 3.8B), demonstrating a normal distribution with little or no bias towards either low, or high, protein copy number per cell.

Proteins identified in each fraction were analysed against known complex components in the CORUM database. A protein complex was deemed to be present in my dataset if a significant proportion of the components within one elution fraction were identified. A significant proportion was defined as being at least 30% for large complexes, but defaulted to accepting only a higher percentage for smaller complexes made up of 6 or less proteins, using the formula max (0.3,2/n). This SEC approach identified 54% of the complexes described in the CORUM database using the strict inclusion parameters described (Figure 3.8C).

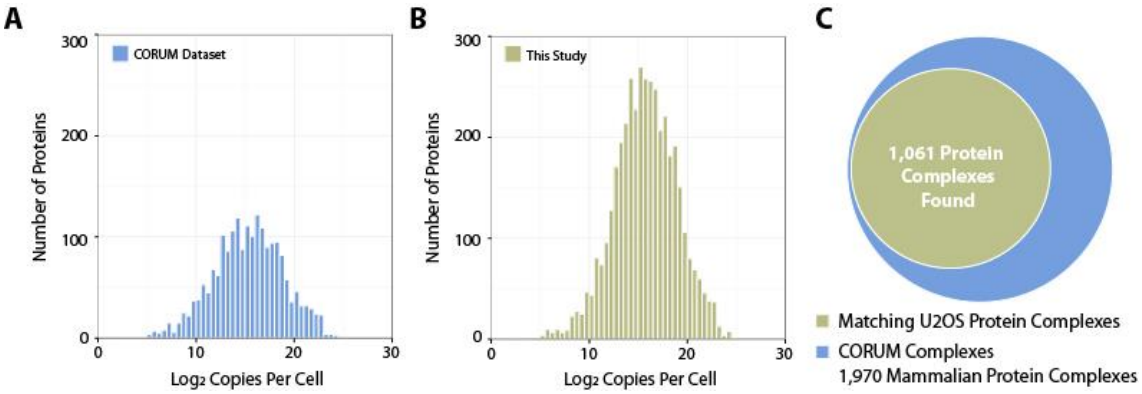


Figure 3.8. Comparison of SEC separated complexes with the CORUM database.

A, protein copy numbers per cell reported in a previous study [Nagaraj et al., 2011] were applied to my SEC dataset. The results are displayed as a histogram comparing the number of proteins (y-axis) and the Log_2 copies of that protein per cell (x-axis). *B*, the protein copy numbers per cell were also applied to the proteins listed in the CORUM database and displayed in the same format as in part *A*. *C*, the proportion of protein complexes found in my SEC dataset that were previously described in the CORUM dataset are presented as a Venn diagram.

3.2.5 Comparative analysis of differentially sized complexes

Subdivision into larger (fractions 7-23), smaller (fractions 24-40), or both size ranges for complex size, revealed them to contain 1,720, 2,280 and 4,076 proteins respectively (Figure 3.9A). Biological process gene ontology (GO) annotations were used to analyse protein function for the dataset [Huang et al., 2008]. The top 10 biological process annotation terms for different size subgroups were identified and this showed that proteins in the overlap group (smaller and larger complexes) are involved in regulatory functions, such as RNA processing and cell cycle regulation (Figure 3.9B). In comparison, proteins in larger complexes were primarily involved in protein transport, catabolic and regulatory processes, such as GTPase activity (Figure 3.9C). I also observed that proteins in smaller complexes were largely involved in phosphorylation and metabolic processes (Figure 3.9D).

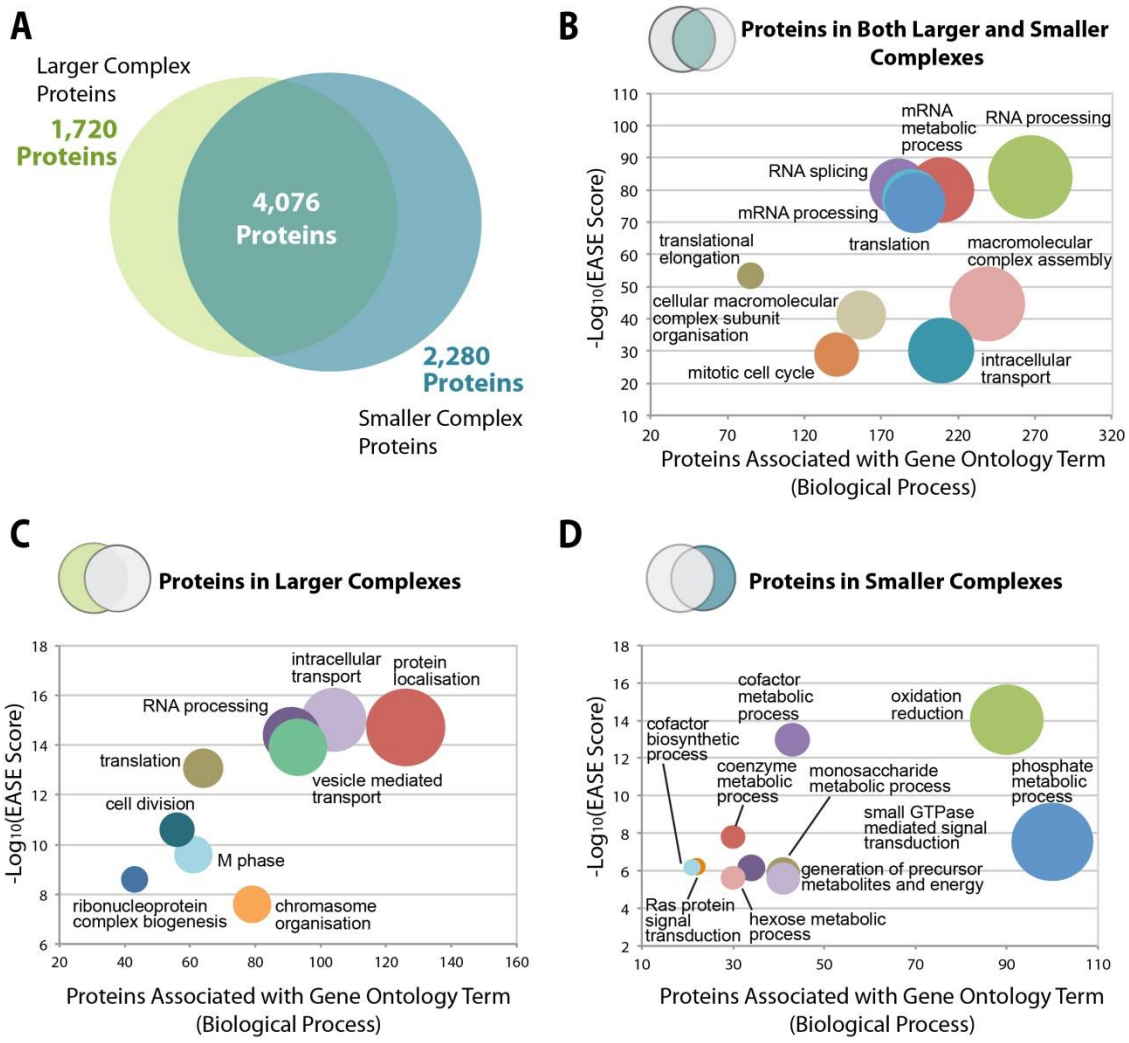


Figure 3.9. Comparative analysis of higher and lower molecular weight complexes.

A, the proportions of proteins identified in the SEC dataset that correspond to "larger complex proteins" (eluting in fractions 7-23) and "smaller complex proteins" (eluting in fractions 24-40) are presented as a Venn diagram. The overlap represents proteins that eluted in both larger and smaller complexes. *B-D*, bubble graphs demonstrating the top 10 Gene Ontology terms (biological processes) (x-axis) plotted against the -Log_{10} (EASE score) for proteins eluting in both larger and smaller complexes, larger complexes and smaller complexes, respectively.

3.2.6 Hierarchical clustering of protein elution profiles

The identification of putative interacting proteins, which should display similar elution profiles, was performed using hierarchical clustering of proteins across the 34 SEC

fractions. To determine the number of protein clusters required for this analysis, 1-1000 clusters were generated using the dataset (Figure 3.10). For each cluster, the correlation coefficient for the proteins within that cluster was calculated to determine optimum cluster size based on similarity of protein profiles. As shown, generating 200 separate clusters for the dataset ensured a high correlation co-efficient, i.e. above 0.9 for each cluster. Choosing a higher number of clusters would overly subdivide protein complexes into separate clusters. I therefore chose to use 200 protein clusters as the optimum number for follow up analysis.

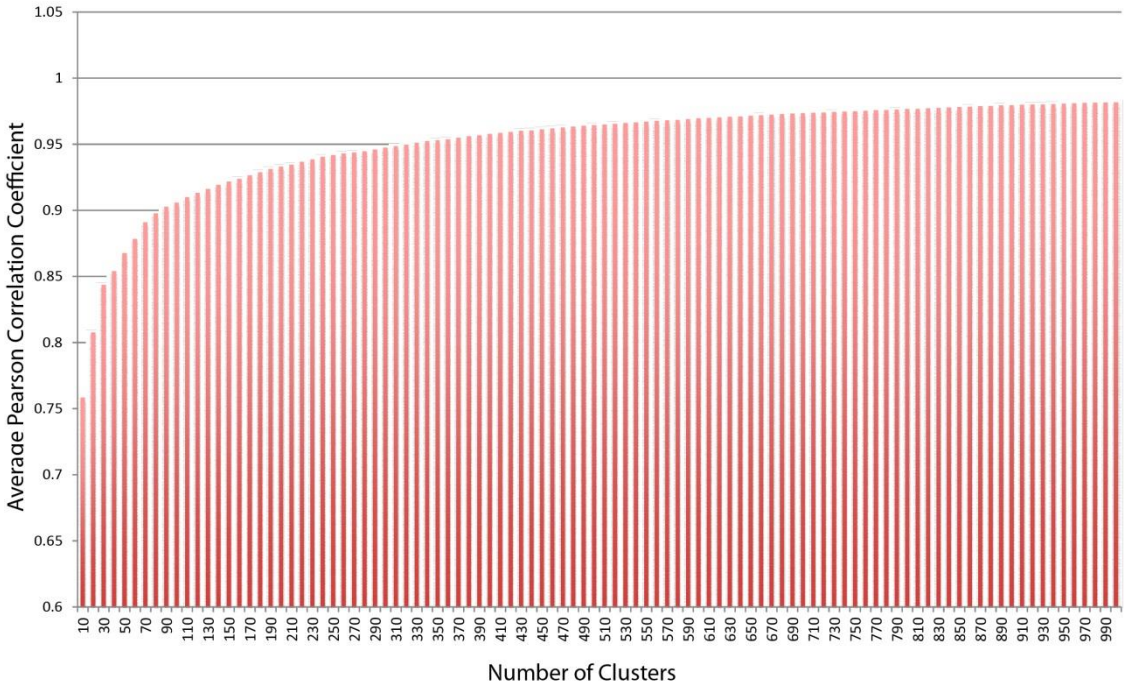


Figure 3.10. Calculation of the optimum hierarchical cluster number.

The column graph shows the average Pearson correlation co-efficient (y-axis) plotted over a range of cluster numbers between 10 and 1000 (x-axis).

Four example protein clusters from the 200 were selected from across the 34 fractions and are indicated on the hierarchical clustering heat map (Figure 3.11, 1-4).

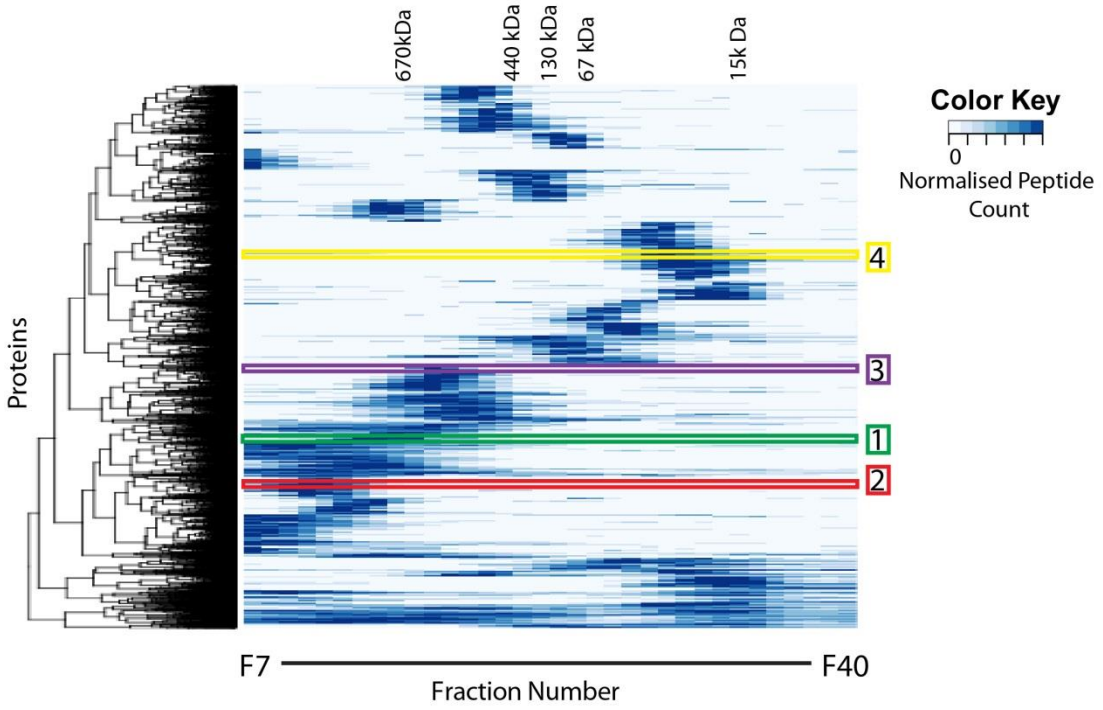


Figure 3.11. Hierarchical clustering of protein elution profiles.

All 8,165 proteins were clustered on the basis of the similarity of their normalised elution profiles across 34 SEC fractions, and the results were presented as a heat map. The dendogram cut-offs for 200 clusters are illustrated to the left of the heat map. The peak elution fraction for each of the protein standards was used to estimate the approximate molecular weight of proteins/complexes eluting within that specific fraction. Four example clusters are highlighted on the heat map to the normalised peptide count.

The components of example clusters were analysed using the STRING database to identify previously known protein-protein interactions within the cluster [Snel et al., 2000; Szklarczyk et al., 2010]. Examination of the components in example cluster 1 (green annotations), showed that it contained many components of the BBS/CCT complex, a hetero-oligomeric complex of about 850 – 900kDa in size [Valpuesta et al., 2002], which correlates to the elution fraction from the SEC column (Figure 3.12A). In addition, interacting elongation factor proteins are also present within this cluster, as

are several lamins (Figure 3.12A). Although these complexes show their strongest elution peak around fraction 14, they can be clearly differentiated using a heat map and the hierarchical tree-based representation of the protein profiles (Figure 3.12B). To further demonstrate this, the elution profiles of each protein within example cluster 1 have been plotted and the average protein profile of each labelled complex has been highlighted (Figure 3.12C). This shows that small differences in the overall elution profile can differentiate these separate complexes.

Analysis of sample cluster 2 contains many components of the (80S) ribosome and elutes from the column at a molecular weight above 670kDa (Figure 3.13). The molecular weight of the intact ribosome is ~4MDa [Verschoor et al., 1996], which is out with the accurate range of the SEC column used. However, I believe that my analysis detected large, intact fragments of the 80S ribosome with an approximate molecular weight of ~1-2MDa, which correlates with a peak in elution fraction 7, as seen for this cluster. Components of the EIF3 complex are also identified within this cluster (Figure 3.13). As the STRING diagram illustrates, there are many established interactions between this complex and the ribosome. This demonstrates that interacting complexes can be identified within the same cluster. Components of the base unit of the 19S proteasome also elute in this same cluster, demonstrating that multiple complexes can exist within one cluster.

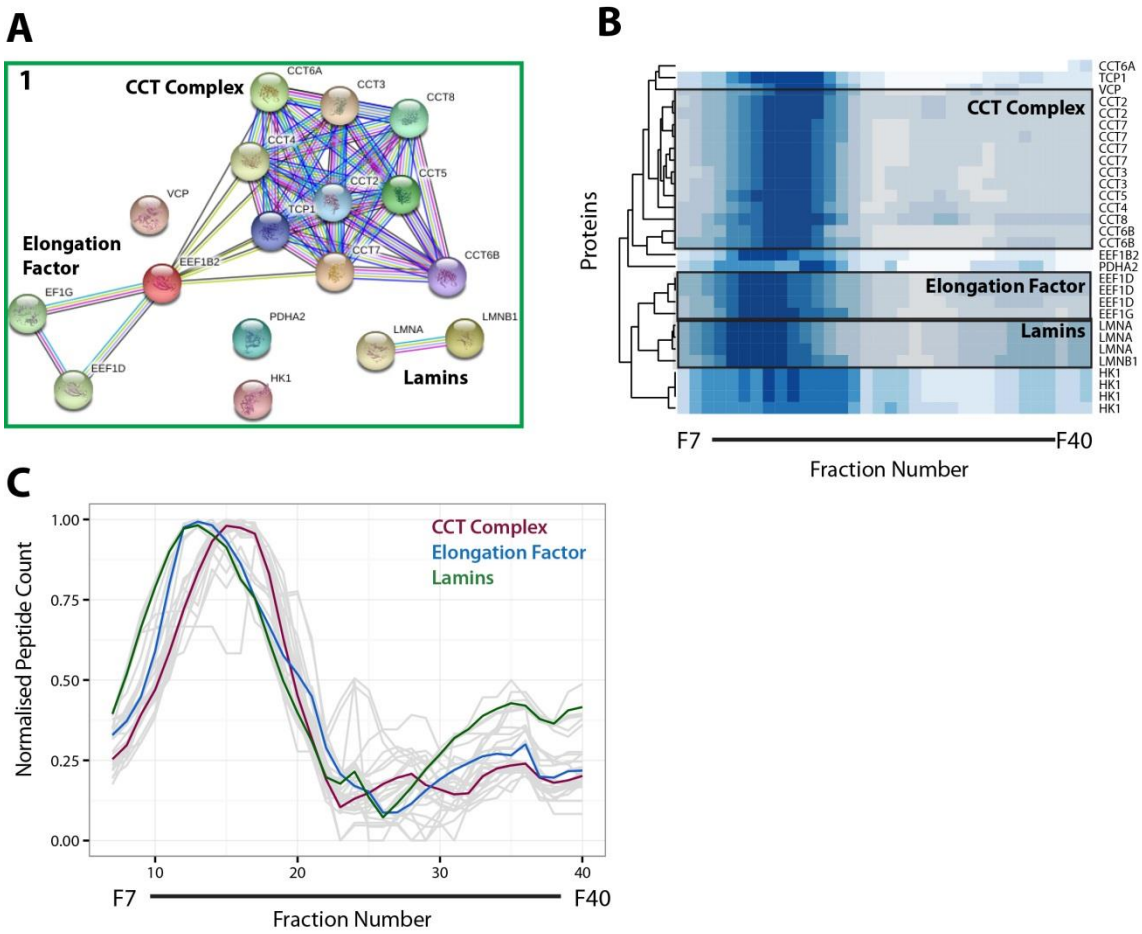


Figure 3.12. Clusters contain proteins known to interact as part of protein complexes.

A, STRING interaction network for proteins identified in example cluster 1 (relating to Figure 3.11). *B*, higher resolution view of the heat map for example cluster 1. Individual complexes previously identified in the literature are highlighted in grey boxes. *C*, the elution profiles for all proteins identified in the example cluster 1 are illustrated as grey lines with their normalised peptide count (y-axis) plotted against the SEC elution fraction number (x-axis). The average elution profiles for co-clustering proteins known to be in highlighted complexes are overlaid in the following colours: chaperonin containing TCP1 complex in red, elongation factor complex proteins in blue, and lamins in green.

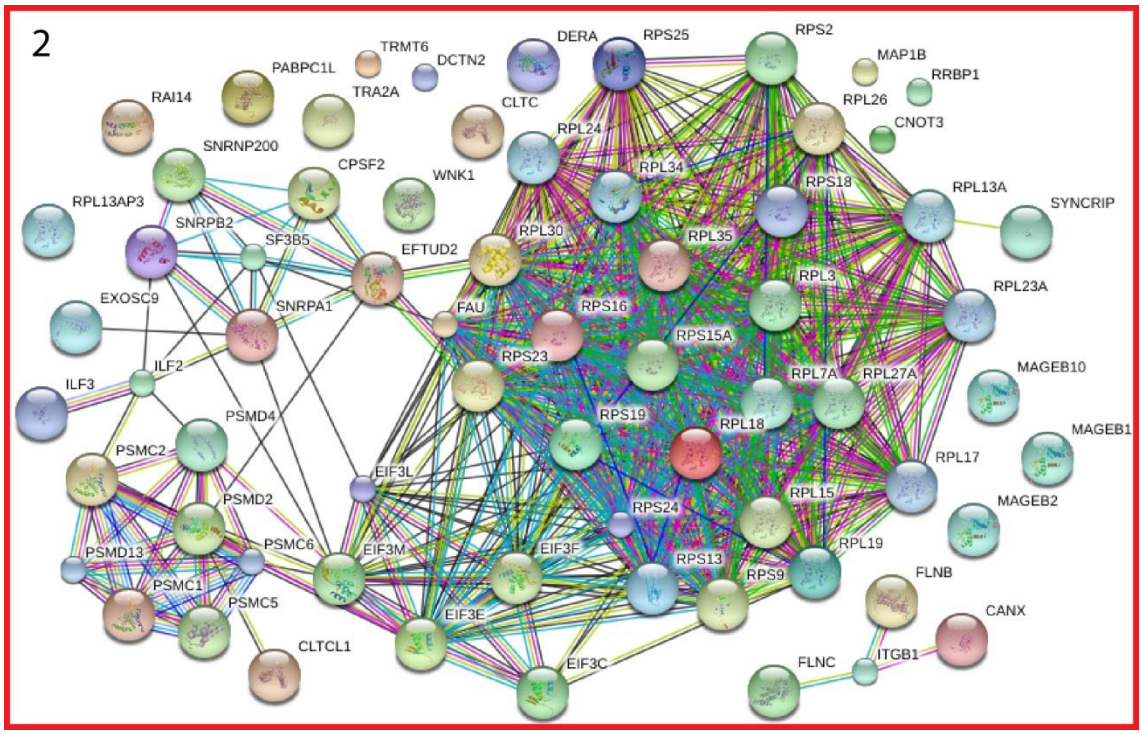


Figure 3.13. STRING diagram for example cluster 2.

STRING diagram for example cluster 2, relating to figure 3.11. Known, or predicted, protein interactions are highlighted by connecting lines between protein names.

Example cluster 3 contained several groups of interacting proteins, including several components of the DNA synthesome complex. This is a multi-subunit complex involved in DNA repair and chromosomal DNA replication [Coll et al., 1996; Jiang et al., 2002] (Figure 3.14, box 3). Example cluster 4 (Figure 3.14, box 4) did not contain large protein complex networks. Within this group, only occasional pairs of interacting proteins were identified, which was expected as the peak elution fraction for this cluster is between 15 and 67kDa.

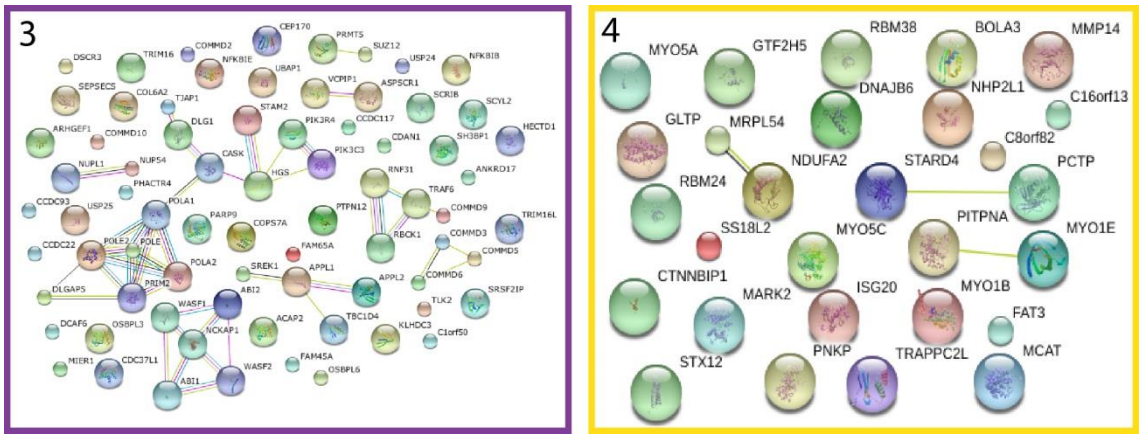


Figure 3.14. STRING diagrams for example clusters 3 and 4.

STRING diagrams for example clusters 3 (purple box) and 4 (yellow box), relating to figure 3.11.

3.2.7 Analysis of protein isoforms

Mass-spectrometry-based analysis enables the unbiased detection of many protein isoforms, such as those generated by either pre-mRNA alternative splice site selection, proteolytic cleavage, or post-translational modifications such as protein phosphorylation. The known isoform annotations for the proteins identified in this study were used to determine whether each isoform of a single protein displayed distinct SEC elution profiles. This identified examples of specific protein isoforms that may participate in different protein complexes.

Alternative splicing

The first example is the alternative splicing of heterochromatin protein 1-binding protein 3 (HP1BP3). This protein is a component of heterochromatin and has been proposed to have a role in modulating chromatin structure and function [Hayashihara et al., 2010]. Four isoforms of this protein are known to be formed by alternative splicing, all of which were identified by up to 14 peptides spanning 26% (isoform 2), to

37% (isoform 5), of the expressed sequences in all four isoforms in this dataset (Figure 3.15A. Peptide locations are marked with an asterisk).

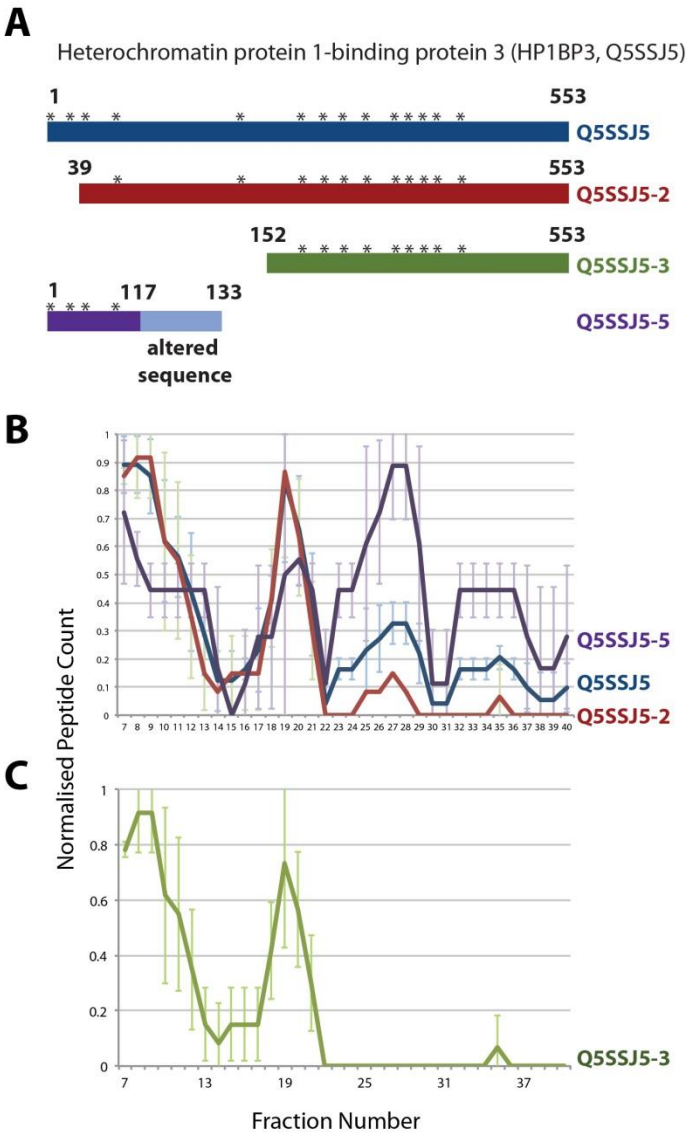


Figure 3.15. Analysis of protein isoforms formed by alternative splicing.

A, schematic of the four isoforms of the HP1BP3 protein. Asterisks indicate the location of identified peptides. SEC elution profiles for the canonical sequence and isoforms 2 and 5 (*B*) and the SEC elution profile for isoform 3 (*C*) are demonstrated as line graphs with the normalised peptide count (y-axis) plotted for each SEC fraction number (x-axis). Error bards denote the standard deviation from the mean.

Closer analysis shows a similar elution profile for three of the four isoforms (1, 2 and 5, Figure 3.15B) whereas isoform 3 is missing a peak between fractions 22 and 30 (Figure 3.15C). As none of the peptides were sequence specific, this could not be attributed to either low expression, or poor sequence coverage.

Although very little is known from the literature about the protein interaction partners of this protein, it could be inferred from the data that HP1BP3 isoforms 1,2 and 5 interact with one or more proteins to form a complex which is in the molecular weight range of ~30-170kDa, whereas isoform 3 does not form part of this complex. Closer analysis of the isoform sequences reveals that a region between amino acids 39 and 116 is missing in isoform 3, but is present in all other isoforms (Figure 3.15A). I hypothesise that this region may be required for an interaction to form the complex(es) that show an elution peak between fractions 22 and 30. This will be an interesting point to pursue in the future as more information regarding this protein, and its isoforms, is uncovered.

Proteolytic cleavage

A second isoform example is demonstrated by the NUP98 polyprotein. The NUP98 mRNA is translated into the NUP98-96 precursor protein, which is then cleaved to form the N-terminal NUP98 protein and the C-terminal NUP96 component as separate proteins (Figure 3.16A) [Fontoura et al., 1999].

The NUP98 pre-mRNA can also undergo alternative splicing to form only the NUP98 protein. The NUP98-96 precursor has six known isoforms and specifically, isoforms 3 and 4 correspond to the NUP98 protein without NUP96. All six isoforms for the NUP98 protein were identified in my dataset with NUP96 containing isoforms (1,2,5&6)

clustering together. These isoforms had a very different elution profile to isoforms 3&4, which co-clustered separately. Figure 3.16B, shows the NUP98 interaction partners identified by STRING.

Figure 3.16C, shows the peptide elution profiles for the isoforms in this dataset. The profiles for isoforms 3&4 (NUP98) are illustrated in orange and the green profile is that of the four isoforms for NUP96. From these profiles I concluded that protein-protein interactions involving NUP96 occurred in fractions 8-17 and those specifically involving the N-terminal region of NUP98 occurred in fractions 22-31. The dataset was interrogated and some of the protein elution profiles for the interaction partners are illustrated in Figure 3.16C. These profiles demonstrate that some interactors have peaks that overlap with the NUP96 component (SEC13, NUP107, NUP133, NUP160, NUP37, NUP43 and SEH1L), whereas RAE1 has a peak corresponding to an interaction involving the N-terminal NUP98 protein. These data suggest that RAE1 may selectively interact with NUP98, but not with NUP96. In addition, these data suggest that the respective NUP96 and NUP98 proteins are involved in distinct protein complexes.

Phosphorylated proteins

Protein phosphorylation is known to affect many protein-protein interactions. The addition of a negative charge can either increase affinity between interaction partners and develop new binding sites, or it can reduce affinity and therefore inhibit interactions [Nishi et al., 2011]. The proteins in my dataset were searched for phosphorylation of serine, threonine or tyrosine residues. The elution profiles of

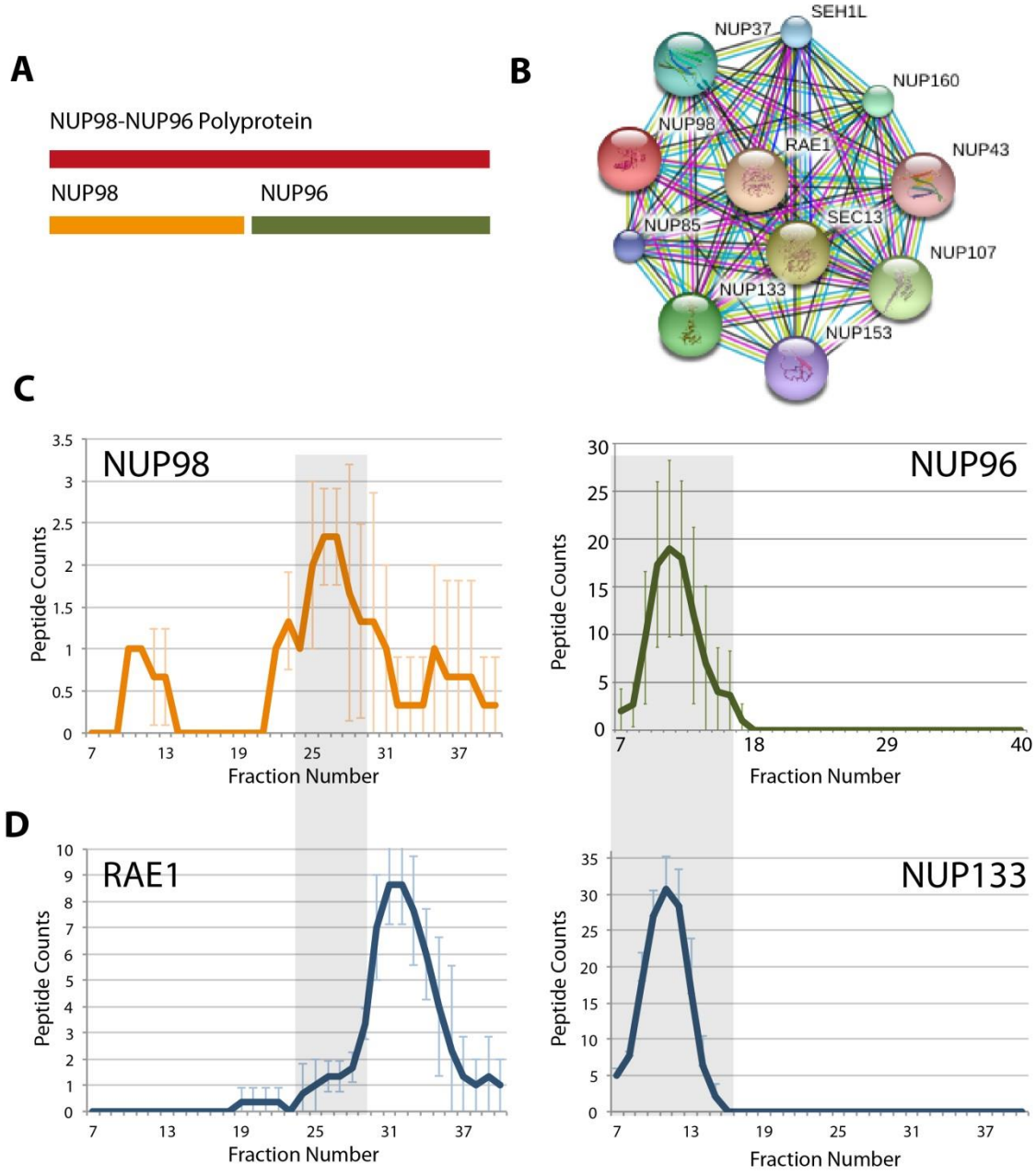


Figure 3.16. Analysis of proteins formed by proteolytic cleavage.

A, schematic showing that the NUP98-NUP96 polyprotein is cleaved to form the separate proteins NUP98 and NUP 96. *B*, NUP98 interacting proteins as proposed by the STRING database. *C*, average SEC elution profiles for peptides associated with specific regions corresponding to the NUP98 protein (orange profile) and the NUP96 protein (green profile). The line graphs represent the peptide count (y-axis) detected for each SEC fraction (x-axis). *D*, SEC elution profiles for RAE1 and NUP133 (x- and y-axes labelled as for part *C*. The grey boxes highlight fractions where I propose RAE1 and NUP98 interact and where NUP133 and NUP96 interact. Error bars indicate standard deviation from the mean.

proteins with and without these modifications were then compared to assess for any discrepancies.

NUDT5 is a protein known to form a homodimer [Zha et al., 2006] and functions in cells to hydrolyse ADP-ribose and other similar molecules such as 8-oxoGDP which can be formed by nucleotide damaging events [Ishibashi et al., 2005]. NUDT5 was detected in this dataset by 13 peptides, representing 64% sequence peptide coverage spanning the full length of the protein (Figure 3.17A). Using this protein as an example, it can be seen that the elution profile for all of the identified peptides contains two main peaks, one in fraction 21 (molecular weight ~440kDa) and the other in fraction 28 (molecular weight ~<67kDa) (Figure 3.17B). Analysis of the dataset for phosphorylated peptides identified phosphorylation of serine 3 in NUDT5 in all three biological replicates. The phosphorylated peptides were all identified in fractions 26-29, corresponding to the second, lower molecular weight peak in the overall protein elution profile (Figure 3.17B,C), whereas the remaining, unmodified peptides, including the peptide containing unmodified serine 3, were identified across the two elution peaks. Interestingly, this phosphorylation site has been identified as an S-Q phosphorylation motif recognised by the DNA damage response kinases ATM and ATR. These data suggest that one or more forms of protein complex containing NUDT5 may be regulated, at least in part, by the phosphorylation of serine 3.

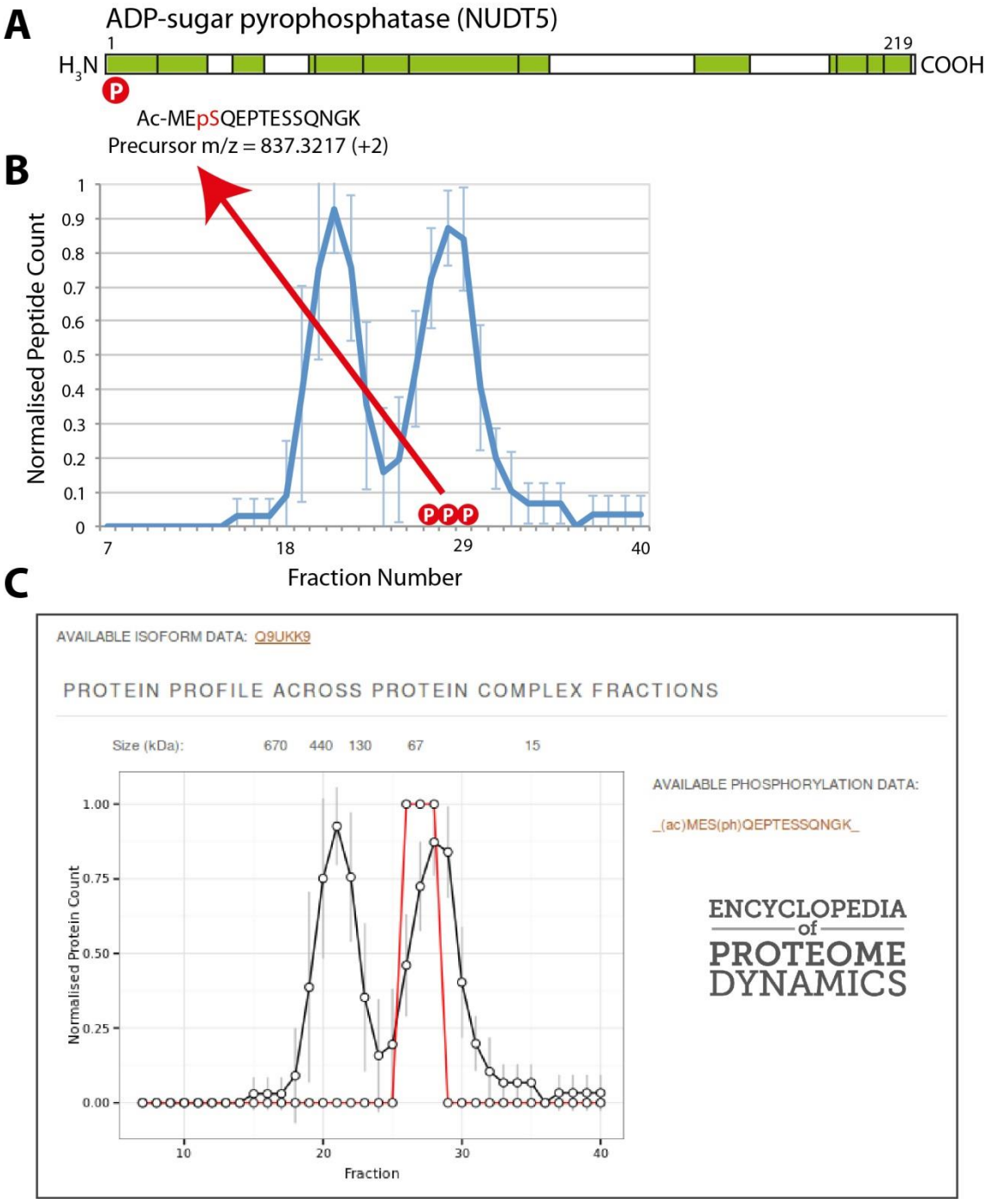


Figure 3.17. Comparison of unmodified and phosphorylated protein profiles.

A, graphical representation of the peptides identified for the NUDT5 protein (shown in green). The letter "P" in the red circle denotes a phosphorylation site on the NUDT5 protein. *B*, the line graph illustrates the SEC elution profile for the NUDT5 unmodified protein. Error bars represent the standard deviation from the mean. The letter "P" in a red circle denotes a fraction containing the phosphorylated peptide (Ac-MEpSQEPTESSQNGK). *C*, a screen shot of these data presented in the Encyclopedia of Proteome Dynamics. Elution profiles show the normalised peptide count on the y-axis plotted against the SEC fraction on the y-axis.

3.3 Discussion

In this chapter, I have used a proteomic strategy combining size exclusion chromatography and mass spectrometry with hierarchical clustering to perform a system-wide survey of native protein complexes in human U2OS osteosarcoma cells. Using this approach, I have been able to allocate interacting proteins into clusters and in addition, I have shown how specific protein isoforms and post-translationally modified forms may interact in different protein complexes. This dataset represents one of the largest experimental surveys of native protein complexes in whole cell extracts, involving the detection and characterisation of the SEC elution profiles for more than 71,600 peptides corresponding to over 8,000 human proteins. The data presented were derived from three biological replicates, which facilitated statistical evaluation in the downstream analysis. This technique demonstrates how the use of a dataset can be expanded beyond simple proteome identification to a more detailed analysis of post-translational modifications and protein isoforms, and to analyse the effects these have on the interactome. The data from this study has been integrated into the Lamond laboratory's 'Encyclopedia of Proteome Dynamics' (EPD) (www.peptracker.com/encyclopediaInformation/), a freely accessible, searchable, online, multi-dimensional data-sharing resource, to give the wider scientific community the opportunity to investigate further any protein or complex of interest.

Hierarchical clustering of protein elution profiles facilitated the identification of protein complex components. The three biological replicates demonstrated that this methodology is reproducible and provide deep coverage of cellular protein complexes. For example, in the native extracts from U2OS cells, I identified more than 50% of known protein complexes published in the CORUM database, which is compiled from a

cumulative analysis of the published literature, including multiple cell types. I have also clearly demonstrated through the use of the STRING database that interacting proteins identified from the published literature were clustered together in this dataset. Additional proteins that co-eluted with known complex components may represent previously unidentified interaction partners, or indeed, may be part of entirely new complexes. Importantly, I have also demonstrated that the combined SEC-MS approach facilitates the identification of protein isoforms, formed either through alternative splicing, proteolytic cleavage and/or by post-translational modifications. I have provided examples of each of these and in each case, I have demonstrated that individual isoforms may be involved in different complexes to the canonical protein. This provides a powerful tool for detecting and annotating protein complexes that augments other techniques commonly used to characterise protein complexes, including affinity pull-down methods and previously described SEC and ion-exchange chromatography methods. I have also demonstrated that the MS-derived elution profiles correlate closely with parallel antibody-based protein detection using western blotting. By incorporating these data into the EPD, the entire dataset is available to the wider scientific community in a searchable form for data mining.

This is one of the most in-depth studies to date of native protein complexes and this is in part due to cumulative technological advances that have been made in mass spectrometry techniques and instrumentation over the last 10 years. These developments have resulted in large increases in the sensitivity and resolution of proteome analyses. However, there is scope for further technical improvements in MS technology that will extend the resolution and sensitivity to detect low abundance proteins and post translational modifications that are under-represented using current

technology. Further developments in software for the analysis of raw MS data will undoubtedly enhance the number of proteins, isoforms and PTMs that can be identified across the SEC fractions. Despite any limitations in resolution capacity associated with SEC, I was able to separate protein complexes at a level that produced biologically relevant and useful information. Advances in SEC column technology, such as modified column designs and modified silica matrices, in addition to the advent of ultra-high performance liquid chromatography (UHPLC) systems, will make this technique even more powerful through increased power of resolution. It is anticipated that such improvements may even facilitate the separation of different complexes that co-elute within one fraction, such as the CCT and elongation factor complexes in Figure 3.12. These developments will also increase the identification of PTMs and thus provide more detail about how the human interactome, and therefore cell function, is regulated by changes in protein complexes.

In comparing different approaches for the analysis of protein complexes, SEC has some clear advantages. One advantage is that the technique can be used with a variety of different buffer conditions. I elected to use a lysis buffer containing physiological levels of sodium chloride to provide insight into the composition of native protein complexes and how they may behave *in vivo*. A different approach is to use Ion Exchange Chromatography (IEX) as per Havugimana *et al.* [Havugimana et al., 2012]. However, this technique requires low salt loading buffer to allow binding of the proteins to the column, which may be detrimental to proteins held together by hydrophobic interactions. None the less, additional SEC analyses under varying salt concentrations and potentially the addition of detergent with SEC could provide valuable information regarding the stability of interactions.

A second benefit of SEC is its relatively high capacity for protein loading. The Superose 6 column I used can be loaded with up to 500µl of sample. The viscosity of the sample, rather than an absolute protein concentration, is a limiting factor for SEC separation but sample concentrations of up to ~70mg/ml may be tolerated. In addition, the SEC approach is highly automated. A highly automated workflow would generate more reliably reproducible data than techniques more reliant on manual procedures, such as BN-PAGE. In this respect my SEC analysis holds a great advantage over BN-PAGE as the SEC column can be integrated with automated HPLC and as a result, protein complexes are separated consistently and remain in a continuous solution from cell lysis through to digestion. Manual loading and cutting of gels required in the BN-PAGE approach introduces variability in the technique, in addition to the risk of sample loss.

I am confident that future applications of the SEC approach will provide a framework for the system-wide analysis of protein complexes under varying conditions. For example, this might include the analysis of complexes in response to different drugs, or cell cycle stages. Provisional results also suggest that the SEC approach described in this chapter can be applied to tissues (Appendix I). Large-scale, high-throughput comparative analyses of complexes under these conditions would not be easily achievable using tagged-affinity purification methods. The fusion of tags to proteins can lead to an alteration in protein function and protein over-expression can also have unwanted side effects [Goel et al., 2000; Rumlová et al., 2001]. These complications do not arise with affinity purification using endogenous antibodies but large scale application of this technique can be limited due to the cost of antibody production, antibody specificity and sensitivity, and epitopes overlapping with interaction sites. Whether tagged or un-tagged affinity purification was attempted, it is not only

financial implications that pose a disadvantage; the sheer complexity of the analysis of the many thousands of pull-down experiments would be immense. These complications are avoided by using the SEC approach as in this technique it is untagged, endogenous proteins in complex that are analysed in parallel. The MS data obtained from this SEC analysis also provides details of protein isoforms and PTMs. These variants are often recognised by the same antibody, therefore traditional IP analyses with western blotting may miss these biologically relevant details. In summary, the workflow I describe is eminently suitable for the large scale analysis of native protein complexes in physiological conditions without the expense and limitations of multiple antibodies and in as automated approach as possible.

Through SEC-MS analysis, three different types of protein isoform were found to form isoform-specific interactions in distinct complexes. One example of this included the NUDT5 protein and its phosphorylated variant. NUDT5 is a nucleotide pyrophosphatase, which is known to catalyse the hydrolysis of a variety of substrates including ADP-ribose, ADP-mannose, ADP-glucose, 8-oxo-dGDP and 8-oxo-GDP. The substrates 8-oxo-GDP and 8-oxo-dGDP are produced following the action of reactive oxygen species on nucleotides. Without the action of hydrolysing enzymes such as NUDT5, the oxidised nucleotides can be incorporated into DNA leading to genetic mutations and/or mistranslation of mRNA [Sekiguchi et al., 2013]. NUDT5 is a 25kDa protein that is known to form a homodimer, which corresponds to the protein elution profile peak seen in fraction 28 of the SEC dataset (Figure 3.17B). Analysis of phosphorylated peptides highlighted phosphorylation of serine 3 of NUDT5 in fractions 27-29 only, overlapping with one of the NUDT5 peaks. This phosphorylation site has been identified previously as an ATM/ATR kinase substrate motif in a large-scale study

using Phospho-ATM/ATR substrate (S*Q) immunoaffinity beads [Hornbeck et al., 2012]. Therefore, I propose that phosphorylation of NUDT5 may be required to mediate the hydrolysis of nucleotides. In the unmodified elution profile, NUDT5 also has a second peak, centred on fraction 21, which corresponds to an approximate molecular weight of ~400kDa. I hypothesise that this peak corresponds to larger multimers of NUDT5 and/or NUDT5 in complex with additional partner proteins/substrates.

Analysis of the SEC data also highlighted the importance of high peptide sequence coverage when identifying proteins. High sequence coverage enhances the ability to discriminate between different protein isoforms [Ahmad et al., 2012]. This was illustrated by the HP1BP3 protein and its isoform variants in my dataset. No isoform specific peptides were identified but the high level of sequence coverage for each isoform facilitated the calculation of an elution profile for each isoform. On the basis of these results, I propose that the protein sequence between positions 39 and 116 may have a role in forming some HP1BP3 protein interactions. HP1BP3 isoform 3, which does not contain this sequence, does not have an elution peak in this region, in contrast to all the other isoforms. Although very little is known about this protein at present, it will be of great interest to investigate further the interaction partners of the different isoforms as more information about this protein is discovered in the future.

I was also able to detect protein isoforms formed by proteolytic cleavage in my dataset. The NUP98-96 polyprotein illustrates how these isoforms may also be annotated as parts of distinct complexes. The nuclear pore complex (NPC) is an extremely large complex comprising approximately 30 different proteins and is highly

conserved from yeast to human [Suntharalingam and Wente, 2003]. The molecular mass of the NPC has been determined to be ~60MDa in vertebrates [Cronshaw et al., 2002]. Therefore, due to the capacity of the SEC column and lack of detergent in the lysis buffer, the intact NPC, if identified, would elute in the void fractions in this study. However, the analysis of NUP98 protein isoforms derived from autoproteolytic protein cleavage of the NUP98-96 precursor has enabled me to differentiate the interactions and probable complex fractions involving the different isoforms. NUP96, the C-terminal component of the precursor polyprotein, corresponds to isoforms 1, 2, 5&6 of the NUP98-96 protein and is involved in the NUP107-160 complex with NUP107, NUP133, SEC13, SEH1L, NUP43, NUP47, NUP85 and NUP160 [Fontoura et al., 1999; Siniossoglou et al., 2000; Belgareh et al., 2001; Vasu et al., 2001; Lutzmann et al., 2002; Harel et al., 2003; Loïodice et al., 2004]. This complex is involved in mRNA export and in the assembly of the NPC itself [Siniossoglou et al., 2000]. In yeast, the NUP84 complex (homologous to the human NUP107-160 complex), has a mass of 600-700kDa as determined by gel filtration [Siniossoglou et al., 2000; Lutzmann et al., 2002; Loïodice et al., 2004]. The Y shaped structure of this complex may cause gel filtration to over-estimate the mass as ultracentrifugation suggests it has a mass nearer to ~375kDa [Siniossoglou et al., 2000]. For proteins in the NUP107-160 complex identified in this dataset, including NUP96, there is a peak in fraction 11, corresponding to a molecular weight >670kDa. I propose that this represents the intact, soluble NUP107-160 subcomplex as it elutes from the column. Although there is a small peak in the NUP98 profile in this fraction, which may represent an intact NUP98-NUP96 interaction [Hodel et al., 2002], the overall profile is very different to NUP96. A larger peak is noted in the NUP98 profile in fraction 26. A similar peak in this fraction is

identified in the profile of RAE1, a protein known to interact with the GLEBS domain in the N-terminal region of NUP98 [Pritchard et al., 1999; Ren et al., 2010]. I hypothesise that this peak represents these proteins interacting as part of a fragment of the NPC as it breaks apart in the SEC column. The larger peak in the RAE1 profile in fraction 31 is most likely to represent monomeric RAE1 (molecular weight 40kDa). As NUP98 is formed by alternative splicing and proteolytic cleavage, it is predicted that it would be more abundant than NUP96 in the cell, but this is not reflected in the profile traces illustrated in Figure 3.16C. One reason for this is that NUP98 is present on the nucleoplasmic and cytoplasmic sides of the NPC and in the nuclear basket [Radu et al., 1995; Griffis et al., 2002; Griffis et al., 2003]. As a result, using a PBS lysis buffer with sonication only may not extract all NUP98, with a significant proportion retained in the pellet. Another possibility is that a significant proportion of NUP98 may also be bound to the intact NPC and therefore elutes in the void. The most striking point is that the profiles for the NUP98 and NUP96 isoforms are different and illustrates once again how the SEC-MS technique can be used to differentiate protein isoform interactions.

The analysis of protein isoforms would not be possible if all identified peptides were attributed to a single gene. Analysing the data at the evidence level in addition to analysing the profiles of phosphorylated peptides in comparison to non-modified peptides, provides a vast amount of biologically relevant data, which may also help define protein complex interactions and functions in the future.

For the data analysis, hierarchical clustering was selected as a method to group proteins with similar elution profiles, and therefore potentially interacting proteins, together. These clusters were formed on the similarity of peptide elution profiles

obtained from the cell lysate across forty fractions. It is possible that collecting a larger number of fractions would increase the resolution of the elution profiles obtained from the lysate, thereby increasing the probability that proteins co-eluting are interaction partners. However, this benefit is offset by the increase in MS run time required to analyse these samples. Increasing the number of clusters may be an alternative approach to increase the confidence that co-clustering proteins interact. However, there is the significant risk that this would result in losing interaction data, particularly regarding intact protein complexes, as these complexes may be fragmented in the clustering process. One possible approach to consider in the future would be to combine the results from two different approaches, for example combining SEC with ion exchange chromatography, and clustering the combined results to increase the confidence in the interactions/resolution.

One limitation of hierarchical clustering relates to proteins that interact as part of multiple different complexes. These proteins are predicted to have multiple peaks across their elution profile, each peak corresponding to a different complex. Unless a potential interaction partner is also a component of the same complexes, the elution profiles of the interaction partners will not match and therefore the proteins will cluster differently. For known interaction partners, this dataset is still of valid use as the individual elution profiles can be compared to look for peak overlaps and thereby predict which complex the interaction belongs to. It is envisaged that the development of the clustering analysis in the future to accommodate individual peak detection will improve the annotation of these multi-complex proteins.

One of the very interesting results highlighted by this study is that there were differences in biological function, as indicated by GO term analysis, between proteins in larger and smaller complexes. This may reflect regulatory differences between these sets of processes. The finding that smaller complexes are enriched for terms associated with metabolic processes may reflect that these proteins/complexes require fewer subunits for regulation and function than the larger complexes. For example, many metabolic reactions may generate a single product from a specific substrate, with the modification being made to the substrate often by one enzyme, albeit with multiple subunits. The resulting product can move by diffusion, without a requirement for any additional protein-protein interactions. In contrast, larger complexes were enriched for terms associated with RNA processing, protein localisation and intracellular transport. These mechanistically complex biological processes may require many proteins to work together as part of large, multi-subunit complexes, such as RNA polymerases and the exocyst and COG complexes. Large complexes, such as the CUL3 E3 ligase complex, may also rely on the interaction of additional protein components to mediate regulation of function in response to temporal and/or spatial cues. In contrast, smaller enzymatic complexes involved in intermediary metabolism, such as dimeric GAPDH, can have their activity controlled by allosteric regulation. As a result, this may require fewer, if any additional protein-protein interactions to facilitate substrate recognition or regulatory functions and hence integrate their activity with metabolic homeostasis.

In summary, this study provides a resource identifying soluble, native protein complexes in U2OS cells, separated on the basis of size. I have grouped proteins with similar elution profiles into 200 clusters, thereby identifying well defined complexes in addition to identifying novel protein interaction networks. The data derived from these

experiments have been incorporated into a searchable, freely accessible, online encyclopedia to provide a useful resource for the wider scientific community. I have demonstrated how these proteomic data can be interrogated to compare complex formation between different protein isoforms formed by splice variation, proteolytic cleavage and post-translational modification. The workflow I have developed lends itself to minor adaptations that would not only facilitate the analysis of native protein complexes in different subcellular fractions, but also to study the response of the native complexes to different salt concentrations, treatment with pharmacological agents including chemotherapeutic drugs, and other stress inducing conditions. Adapting the workflow to incorporate pulse labelling by SILAC is also a very exciting prospect to consider for future analysis as this would facilitate a greater understanding of the order of complex formation. All of these will be interesting avenues for future research, but the development of this technique to study native protein complexes in animal and human tissues would be ground breaking in the field of interactome analysis.

3.4 Distribution of Effort

Dr Yasmeen Ahmad performed the R-based statistical analysis of the SEC data, including the analysis of protein abundance in this dataset and the CORUM database, and the gene ontology analysis although I assisted in planning parameters for statistical analysis. Dr Mark Larance ran the samples through the Superose 6 SEC column and the peptides were analysed on the LTQ Orbitrap Velos mass spectrometer in the "FingerPrints" proteomics facility at the College of Life Sciences, University of Dundee. Data derived from this study was published in 2013 [Kirkwood et al., 2013] (Appendix II).

Chapter 4: Protein crosslinking and endogenous complex analysis

4.1 Introduction

As discussed in the previous chapter, few proteins function alone but instead most function as part of protein complexes to carry out cellular functions. Some protein-protein interactions are transient, such as reversible substrate-enzyme binding, and others are very weak, yet still vital to many biological processes. These transient and weak interactions are likely to be amongst those not identified in the study of native complexes in chapter 3. I believed that the addition of a crosslinker would facilitate the preservation and subsequent identification of these interactions, thereby improving resolution and depth of coverage of endogenous protein complexes. It is assumed that proteins interacting with each other, either directly, or as part of a larger complex, are likely to lie in close proximity to one another. Therefore, the addition of a reagent that holds these proteins together throughout cell lysis would enable more efficient recovery and detection of these complexes. The use of a crosslinker should also reduce non-specific interactions, which might arise from cell lysis when proteins mix with those from other subcellular compartments, providing a more accurate analysis of endogenous protein complexes. In addition, crosslinking proteins enables the use of denaturing lysis buffer conditions, thereby increasing the extraction of complexes tightly bound to cell substructures, such as membrane proteins, without disrupting protein-protein interactions.

4.1.1 Types of crosslinkers

Protein crosslinking refers to the formation of a covalent bond either within (intramolecular), or between, molecules (intermolecular). Crosslinking compounds are generally bifunctional; meaning that they have two reactive sites separated by a spacer region. The chemical reactivity of the crosslinker will determine its suitability for different uses. The length of the spacer/linker portion will also affect the distance between sites that can be crosslinked. Several different crosslinkers exist and the most commonly used types are listed in the following section defined by their reactive sites [Sinz, 2006].

- Amine-reactive crosslinkers (NHS esters, Imidoesters, Carbodiimides)
- Sulphydryl-reactive crosslinkers (maleimides)
- Photoreactive crosslinkers (aryl azides, diazirines, benzophenones)

Crosslinkers are generally divided into either homo-bifunctional, or hetero-bifunctional agents. Homo-bifunctional crosslinkers have two identical reactive sites (e.g. amine reactive-amine reactive, including dithiobis[succinimidyl propionate] (DSP) and Bis[sulfosuccinimidyl] suberate (BS3), whereas hetero-bifunctional agents have different reactive sites at either end (e.g. sulphydryl reactive-amine reactive; examples include succinimidyl-4-[N-maleimidomethyl] cyclohexane-1-carboxylate (SMCC) and N-succinimidyl 4-[4-maleimidophenyl]butyrate (SMPB).

The choice of crosslinker depends on the proposed experiment, and factors such as membrane permeability, reversibility and spacer arm length should all be considered. For topological protein complex analysis using crosslinking and MS, BS3 [Chen et al., 2010; Kalisman et al., 2012; Robinson et al., 2012; Murakami et al., 2013], DSS [Lasker

et al., 2012; Leitner et al., 2012; Nguyen et al., 2013; Tosi et al., 2013] and SMCC have been utilised [Rappsilber et al., 2000]. Reversible crosslinkers, such as DSP [Kim et al., 2002; Meunier et al., 2002; Kim et al., 2006], and formaldehyde [Vasilescu et al., 2004; Tagwerker et al., 2006], have also been used to stabilise protein interactions that would otherwise be lost upon cell lysis in affinity purification studies. DTBP, a membrane permeable crosslinker, has been used to stabilise integrin complexes [Humphries et al., 2009; Byron et al., 2012]. Without membrane-permeable crosslinking, known integrin-complex components cannot be immunoprecipitated [Humphries et al., 2009].

4.1.2 Use of crosslinking in protein complex analysis

Structural protein complex analysis

Rappsilber *et al.* first combined protein crosslinking with mass spectrometry to identify components of the NUP84 complex [Rappsilber et al., 2000]. Since then protein crosslinking has also been used for the targeted study of such complexes as the 19S proteasome lid [Sharon et al., 2006], the 26S proteasome [Lasker et al., 2012], the Ndc80 complex [Maiolica et al., 2007], GRP94 [Chu et al., 2006], RNA polymerase II complexes [Chen et al., 2010; Murakami et al., 2013], the TRiC/CCT subunits [Kalisman et al., 2012; Leitner et al., 2012], modules of the transcription factor, Mediator [Robinson et al., 2012; Larivière et al., 2013], the ribosomal protein S1 [Lauber et al., 2012], and the SWR1 [Nguyen et al., 2013] and INO80 [Tosi et al., 2013] chromatin remodelling complexes. These studies all performed crosslinking on purified protein complexes and analysed the crosslinked peptides by mass spectrometry. The MS results of the crosslinked peptides are cross referenced to a library of all the possible peptide modifications for the proteins of interest, which enables the calculation of the

complex structure. In general, purification of the complex is deemed necessary, as without doing so, the reference library would be too large. Yang *et al.* suggest that purification of complexes is not essential in crosslinking studies, and propose the use of their software program, pLink, as a solution for the analysis of crosslinked peptides [Yang et al., 2012]. This software is reported to detect regular peptides, mono-linked (crosslinker binds to one reactive group only with the other active site deactivated), loop-linked (intramolecular crosslinks) and interlinked peptides (links between two peptides). Utilising this software, this group identified 394 interlinks in BS3-treated *E.coli* lysates and 39 interlinked peptides in *C.elegans* [Yang et al., 2012]. Only two previous crosslinking studies had attempted crosslinking using cell lysates, both in *E.coli*, without identifying a higher number of interlinks [Rinner et al., 2008; Xu et al., 2010]. This illustrates that crosslinking analysis of intact protein complexes for structural analysis on a large scale has only been performed to a limited extent.

Crosslinking in affinity purifications

Many protein-protein interactions are either transient, or weak, and as a result, cannot be detected under normal affinity purification conditions. To counteract this, cell permeable crosslinking agents have been added either to live cells [Meunier et al., 2002; Vasilescu et al., 2004; Tagwerker et al., 2006; Humphries et al., 2009; Byron et al., 2012], or to cell lysates [Kim et al., 2002], to stabilise these interactions. Examples where such approaches have facilitated the identification of complexes include: raptor as an interaction partner of mTOR with a role in nutrient-stimulated signalling and regulation of mTOR kinase activity [Kim et al., 2002]; RCC2, a component of an integrin-associated complex, as a regulator of Rac1 and Arf6 [Humphries et al., 2009]; α-subunit-dependent protein recruitment in integrin adhesion complexes [Byron et al.,

2012]; molecular chaperones as part of a large endoplasmic reticulum-localised mulitprotein complex [Meunier et al., 2002]; keratin 17 as an interactor of 14-3-3 to regulate cell growth [Kim et al., 2006] and IQGAP1 as an interaction partner of M-Ras^{Q71L} [Vasilescu et al., 2004].

4.1.3 Formaldehyde and the analysis of protein interactions

Formaldehyde is a homo-bifunctional crosslinking agent that has been utilised for centuries as a fixative to preserve tissue architecture. Due to its small size, formaldehyde readily permeates the cell membrane, making it ideal for the study of protein interactions in intact cells and tissues. This enables the fixation and crosslinking of proteins, DNA and other reactive molecules in both time and space, through the formation of covalent bonds. As a result, it has widely been used for the study of such features as cellular architecture in immunohistochemistry; protein-DNA interactions in chromatin immunoprecipitation (ChIP) assays [Kuo and Allis, 1999; Orlando, 2000]; protein-protein interactions by means of single [Vasilescu et al., 2004] and tandem affinity purification [Guerrero et al., 2006; Tagwerker et al., 2006], endogenous protein purification [Bai et al., 2008], in isotope labelling methods [Guerrero et al., 2006; Bai et al., 2008], non-denaturing [Schmitt-Ulms et al., 2004], and denaturing purification conditions [Guerrero et al., 2006; Tagwerker et al., 2006], in cell culture and in tissues [Schmitt-Ulms et al., 2004; Bai et al., 2008]. The specific crosslinking sites in the above experiments were not determined.

The chemistry of formaldehyde crosslinking is presumed to be more complex than that of other crosslinking agents, and the resulting peptides to be much more heterogeneous and difficult to assess by mass spectrometry [Toews et al., 2008]. Studies investigating formaldehyde crosslink modifications have shown modifications

of many different amino acid residues following extended formaldehyde exposure, over several days [Metz et al., 2004; Metz et al., 2006], which seemingly confirms the complexity. However, Toews *et al.* have claimed that short term exposure to formaldehyde, i.e. 10-20 minutes, led to negligible peptide crosslinking [Toews et al., 2008]. Thirty Dalton increases in peptide mass were only observed as either amino group intermediates, or as thiohemiacetal products of cysteine [Toews et al., 2008]. The major reaction products observed in this study were increases in multiples of 12Da associated with peptide amino-termini, lysine and tryptophan residues [Toews et al., 2008].

4.1.4 Formaldehyde chemistry

Formaldehyde reacts with water to form methylene glycol, HO-CH_2-OH which can form polymers, usually containing between 2 to 8 units (Figure 4.1A). Larger polymers of formaldehyde (containing up to 100 subunits) are insoluble in water and termed paraformaldehyde. Paraformaldehyde can be converted, either to formaldehyde monomers, or small polymers, by heating to 60°C in a buffered solution at physiological pH (Figure 4.1A). For use as a fixative, the majority of the resultant solution must comprise monomeric formaldehyde. It is the aldehyde group in formaldehyde (-CHO) that reacts with nitrogen and some other atoms in proteins and if the aldehyde group reacts with two atoms in close proximity a –CH_2 crosslink, (methylene bridge), is formed (Figure 4.1B). As formaldehyde contains only four atoms, the linker formed between proteins is very short, in the region of 2.3-2.7Å. Formaldehyde crosslinks are less specific in their targets than other crosslinkers, with bonds forming between side chains of cysteine, tyrosine, tryptophan, asparagine, glutamine, arginine, histidine, lysine and amino termini. It has been shown that under

short fixation periods (10 minutes) the majority of the crosslinks involve lysine, tryptophan side chains and the amino termini [Toews et al., 2008], however prolonged fixation leads to crosslink formation across many amino acid residues [Metz et al., 2004; Metz et al., 2006].

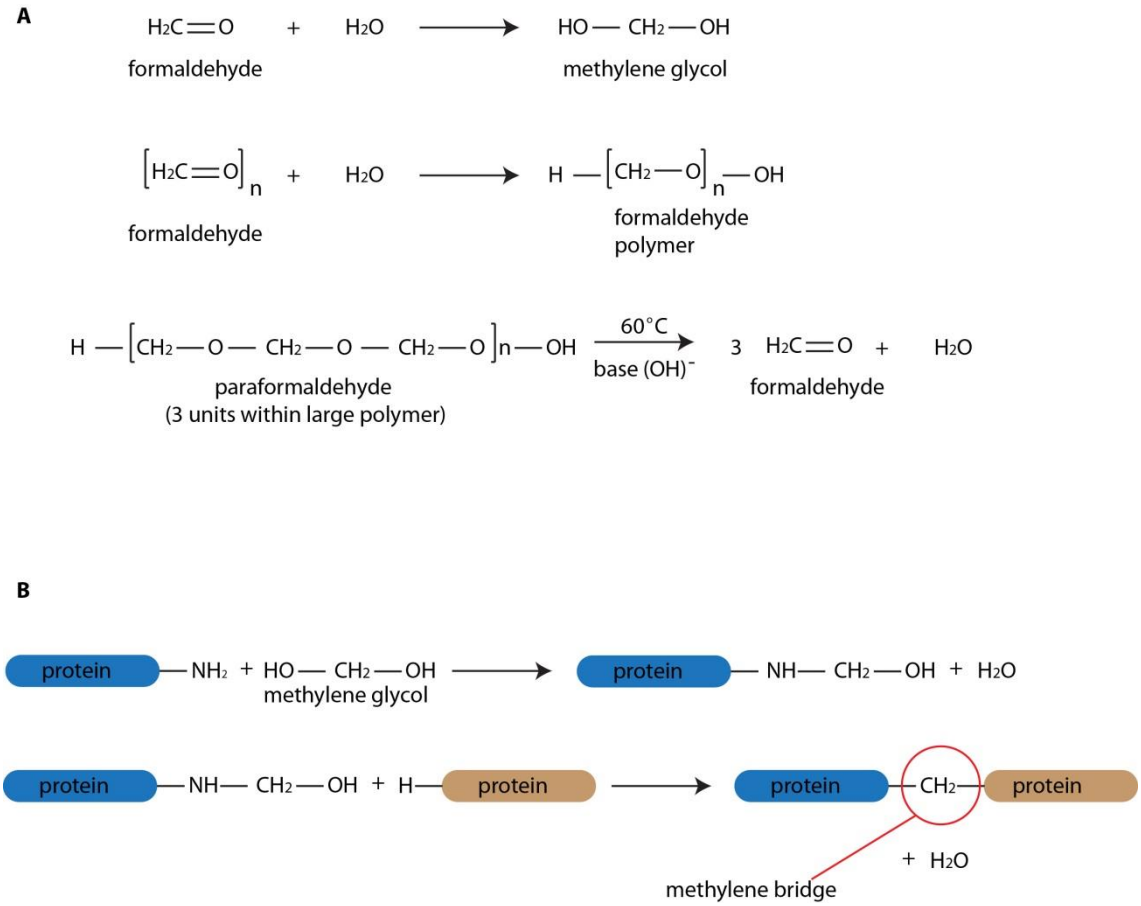


Figure 4.1. Formaldehyde chemistry.

A, formation of methylene glycol, formaldehyde polymers and depolymerisation of paraformaldehyde. *B*, reaction of formaldehyde with proteins.

The ease of formaldehyde crosslink reversal, by heating, is key to the use of this agent as a fixative prior to immunohistochemical analysis, a feature that would also be important for identifying proteins in complexes using SEC. Taking all of these points into account, i.e., ability to permeate cells; rapid onset of crosslink formation; non-

specificity of the target permitting crosslinking of a variety of subunits; compatibility with MS studies; and strong yet reversible bonds, this is the crosslinker I opted to utilise in the first instance. In addition, this also enabled me to use a denaturing buffer to improve protein extraction from membrane and nuclear compartments because the crosslinks prevent SDS breaking the interactions.

4.1.5 Hypothesis and Aims

- Formaldehyde protein crosslinking can be combined with SEC chromatography to enhance recovery of complexes and increase resolution in the study of protein complexes

To do this I have:

- Optimised formaldehyde crosslinking *in vitro* combined with denaturing cell lysis to preserve protein complexes
- Separated crosslinked protein complexes by size using SEC before crosslink reversal and analysis by LC-MS/MS

4.2 Results

4.2.1Workflow for the identification of crosslinked protein complexes

Crosslinking of native proteins was performed *in vivo* with formaldehyde in PBS for 10 minutes before quenching the reaction with glycine. Cells were lysed in a buffer containing 4% SDS, sonicated then incubated at 37^{o}C for 30 minutes to aid denaturation of complexes, whilst maintaining crosslinks. The crosslinked protein complexes were then separated by size/shape using denaturing SEC. For this experiment, a different SEC column was used to that used in chapter 3, the BioBasic SEC1000. This column contains particles smaller in size than the Superose 6 column,

each with 1000Å pores. These features improved the resolution of complex separation, and as a result, co-eluting proteins were more likely to be true interaction partners/components of the same complex. The eluate was collected in 48 equal, sequential fractions; the larger number of fractions was selected as an additional attempt to improve resolution. The fractions were then heated to 95^{o}C for 30 minutes to reverse the crosslinks. For mass spectrometry analysis, eluate fractions were digested with both trypsin and endopetidase lysC, because double protease digest has been reported to improve sequence coverage [Ly et al., 2014]. Peptide identification and data analysis was performed as per native protein complexes described in the previous chapter. The general workflow for crosslinked protein complex analysis is described in Figure 4.2.

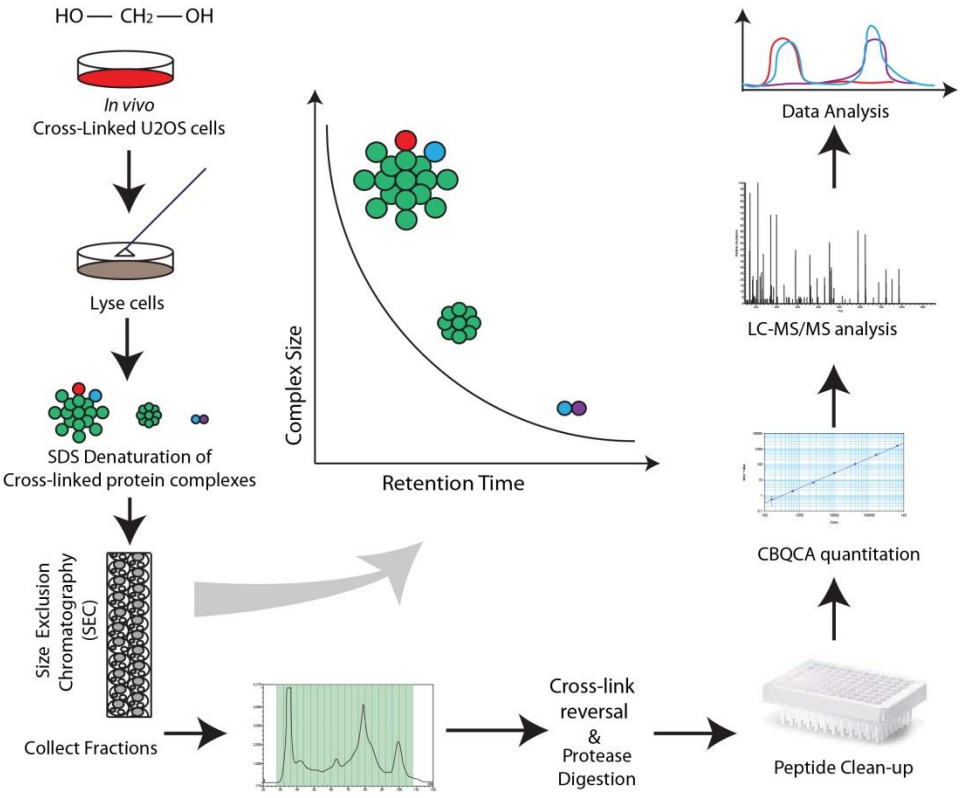


Figure 4.2. Endogenous protein complex analysis using a comprehensive proteomics workflow.

Workflow for SEC-based separation of crosslinked protein complexes.

4.2.2 PFA crosslinking preserves protein complexes through SEC

To assess if crosslinking would hold protein complexes intact through denaturing lysis and SEC separation, U2OS cells were incubated with 1% formaldehyde. This concentration has previously been utilised for Chromatin-immunoprecipitation (ChIP) [Kuo and Allis, 1999] and in other formaldehyde based crosslinking studies [Vasilescu et al., 2004; Guerrero et al., 2006; Tagwerker et al., 2006]. Samples were prepared as described in the workflow above to the stage of crosslink reversal (Figure 4.2). The SEC chromatograms demonstrate that crosslinked proteins started to elute from the column approximately four minutes before non-crosslinked proteins and that the absorbance in these early fractions is higher for crosslinked proteins than non-crosslinked. Conversely, the non-crosslinked protein elution profile showed higher absorbance in the later eluting fractions compared to the crosslinked profile. These results suggested that 1% formaldehyde did crosslink proteins into larger complexes compared to untreated cells (Figure 4.3). Protein standards prepared in the same lysis buffer were used to calculate the approximate molecular weight/size of complexes eluting across the fractions. These show that the BioBasic SEC1000 column can separate protein complexes in the range of 1.8MDa to 15kDa.

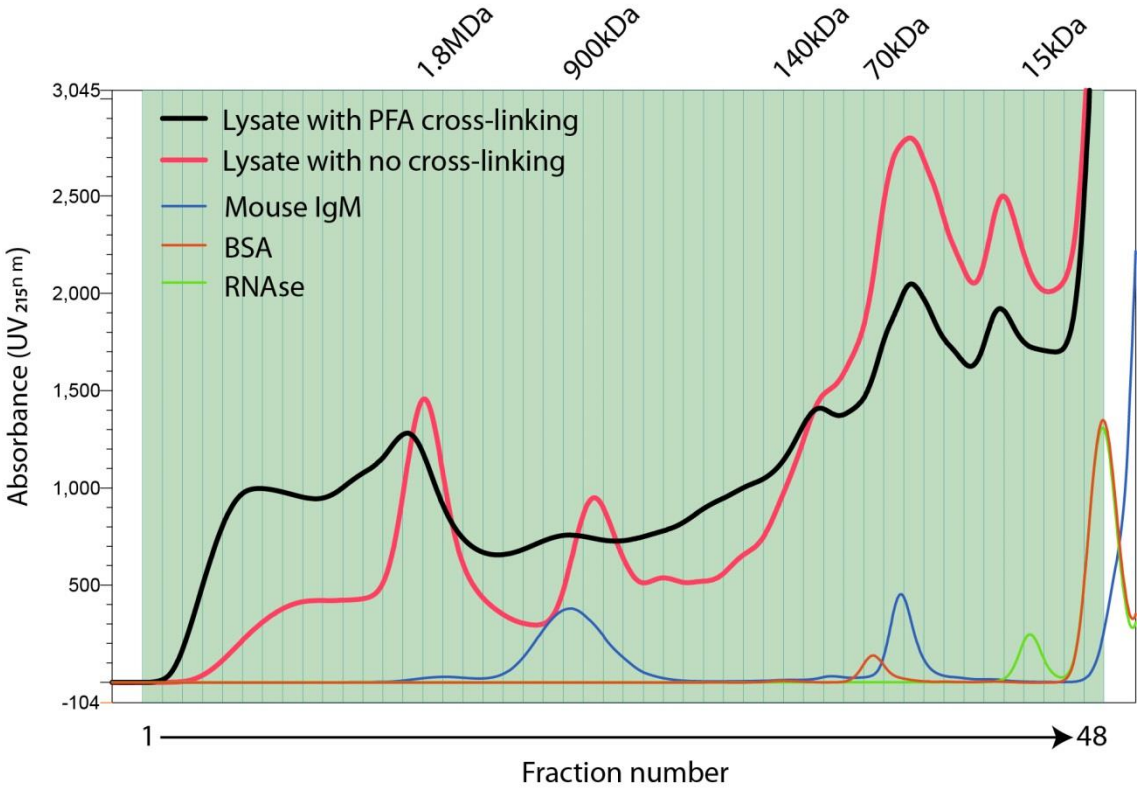


Figure 4.3. Characterisation of protein complex separation.

The UV chromatograph from one of the biological replicates indicates the profile of the U2OS total cell lysate as it eluted from the SEC1000 column across 48 fractions. Protein standards of known molecular weights (IgM, BSA, RNase) were injected onto the same column, and their elution peaks were used to calculate approximate molecular weights for fractions.

After crosslink reversal, the same representative aliquots of the eluates from crosslinked and non-crosslinked lysates were prepared for SDS-PAGE. SYPRO ruby staining of the selected fractions confirmed that early elution fractions without crosslinking contain virtually no detectable protein. However, with crosslinking, the same fractions contain a range of proteins of variable molecular weight (Figure 4.4). These results demonstrate that formaldehyde crosslinks could withstand cell lysis in a denaturing buffer and that these crosslinked complexes could be separated by SEC.

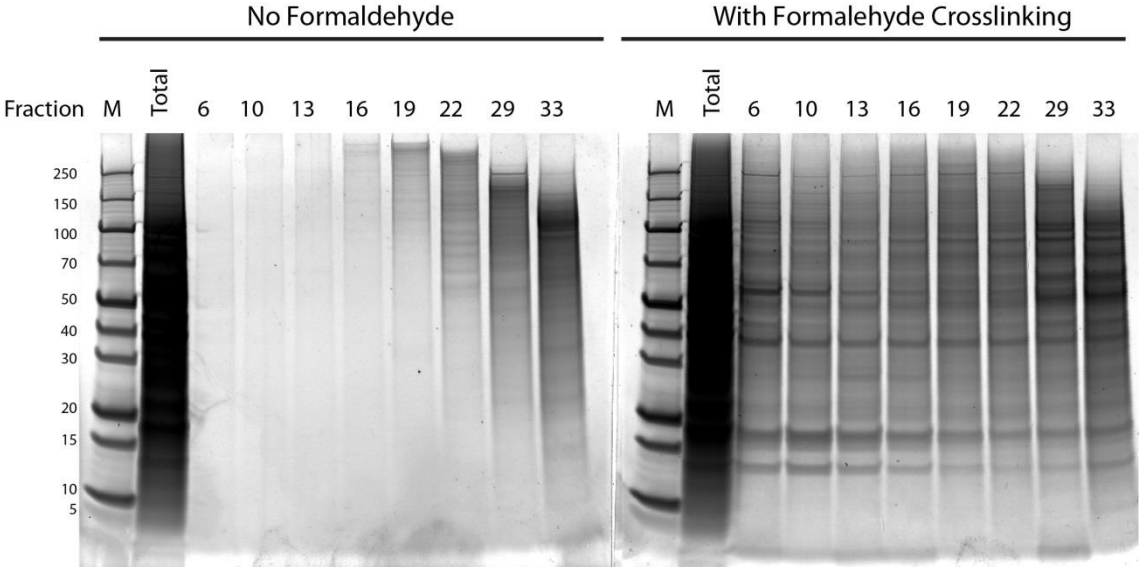


Figure 4.4. Formaldehyde crosslinks preserve protein complexes in denaturing conditions.

SYPRO ruby total protein stain of proteins in the total cell lysate and selected fractions of cells treated without (left panel) and with (right panel) formaldehyde crosslinking. 10μg of protein was loaded in lane 2 (total), and a maximum of 10μg of protein was loaded per lane for elution fractions.

4.2.3 Six percent formaldehyde is the optimal crosslinking concentration for broad range protein analysis combined with SDS lysis

Although 1% formaldehyde was shown to hold some complexes intact, the chromatogram demonstrated the main elution peaks were in low molecular weight fractions, raising the possibility that some complexes were still dissociating under these conditions. To determine the optimal concentration of formaldehyde to use with the SEC analysis, I prepared cells in the presence of 0, 3, 4, 5, 6 or 8% formaldehyde prior to lysis. Equal amounts of protein from each sample were then prepared for SDS-PAGE and immunoblotted for different proteins that are either known to form polymeric chains, or to be part of large protein complexes (Figure 4.5). The results show that under my experimental conditions, significant amounts of the proteins

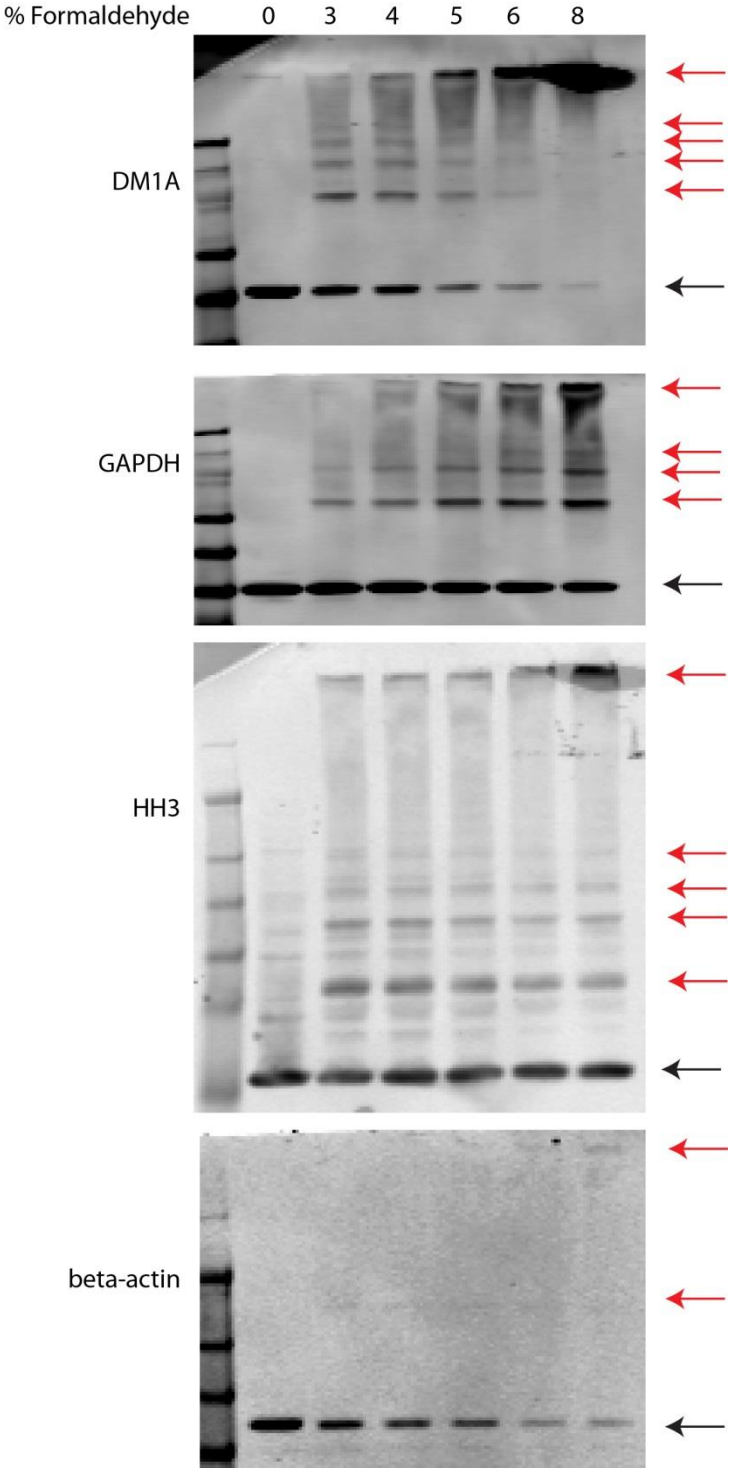


Figure 4.5. 6% formaldehyde is suitable for broad range protein analysis combined with SDS lysis.

Immunoblots of total cell lysates from U2OS cells treated with varying concentrations of formaldehyde. 10μg of total protein loaded per lane. Black arrows indicate monomeric protein and red arrows indicate crosslinked multimers.

remain in monomeric form with less than 6% formaldehyde fixation. Eight percent fixation showed the potential to over crosslink some proteins (Figure 4.5). Therefore 6% formaldehyde was selected as the optimum strength of crosslinker for my future experiments.

Comparison of the chromatograms from cells crosslinked with 1% and 6% formaldehyde show an increase in UV absorbance detected in early fractions in 6% formaldehyde lysates, with a relative decrease in the absorbance detected in later fractions (Figure 4.6). These results show that 6% formaldehyde crosslinks more proteins in larger complexes than 1% formaldehyde under these experimental conditions.

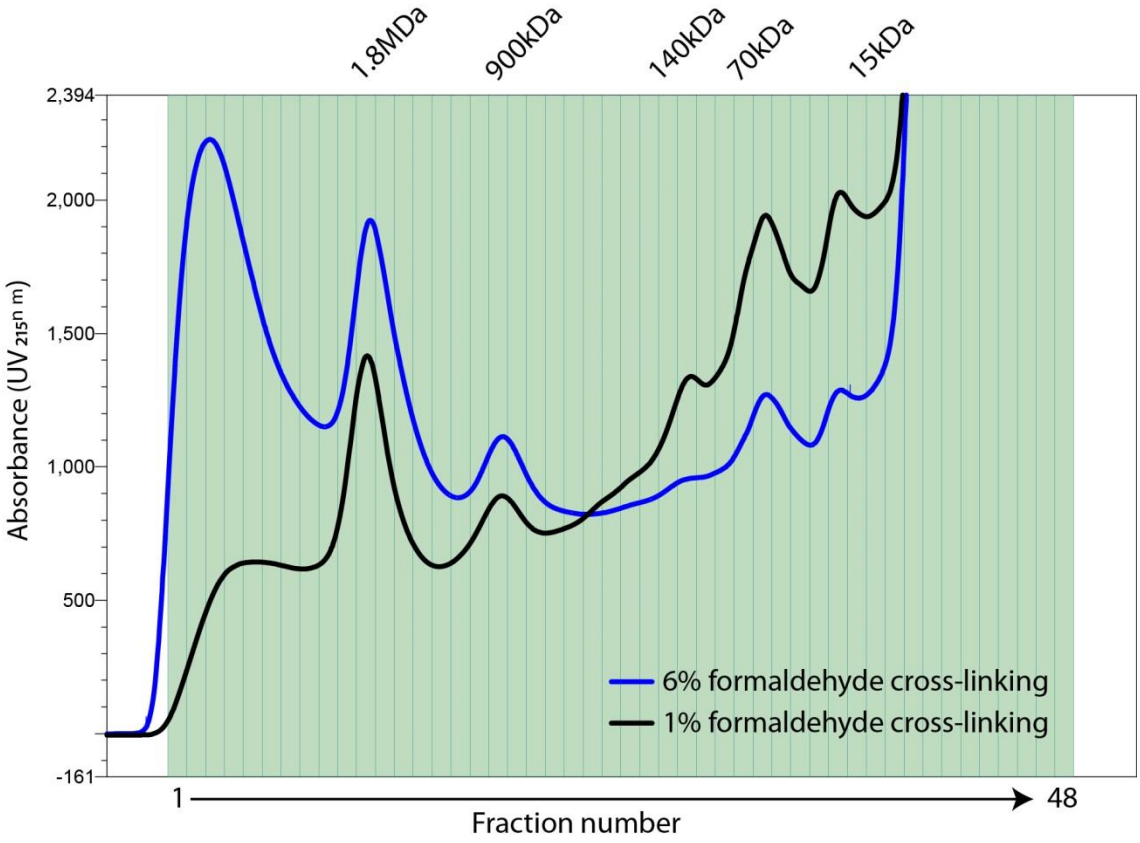


Figure 4.6. Formaldehyde crosslinks proteins in large complexes.

The UV chromatograms for lysates from cells treated with either 1%, or 6%, formaldehyde as they elute from the SEC1000 column across 48 fractions.

SYPRO ruby staining of eluate fractions from protein complexes crosslinked with 6% formaldehyde confirm that the size separation successfully separated complexes from >460kDa to ~14kDa. The estimated size of each fraction based on SEC protein standards closely resembles the molecular weight of crosslinked proteins detected by SYPRO ruby staining as they migrate during gel electrophoresis (Figure 4.7).

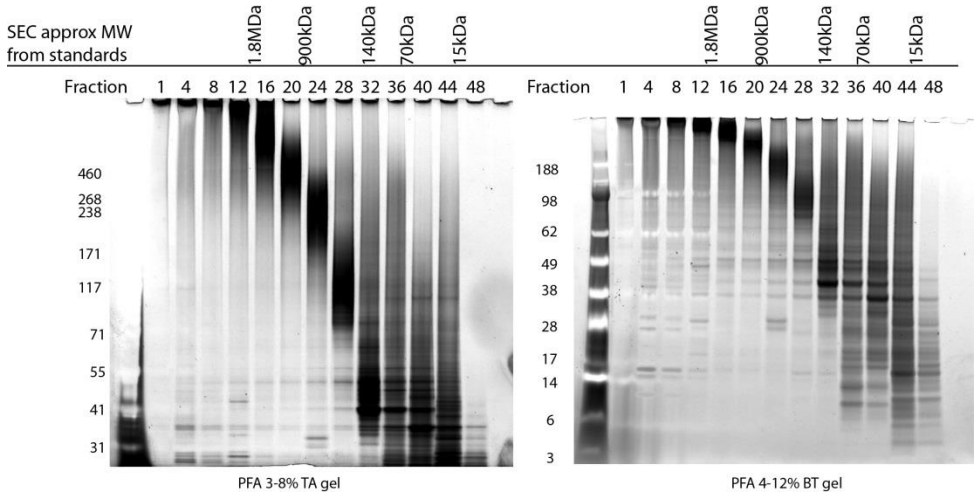


Figure 4.7. Characterisation of protein complexes separated by crosslinking and denaturing SEC.

SYPRO ruby total protein stain of selected fractions run on 3-8% Tris acetate (left panel) and 4-12% Bis-tris gels shows close correlation between SDS-PAGE estimated molecular weights and SEC protein standard molecular weights. A maximum of 10μg of total protein was loaded per lane.

4.2.4 Coverage obtained using crosslinking and denaturing techniques

Using strict inclusion criteria of a minimum identification of two unique peptides in all three biological replicates, I identified a mean of 88,568 peptides, corresponding to 6,425 proteins. Eight hundred and thirty-three of these peptides carried a formaldehyde modification resulting in a 12Da mass increase, and 1,712 carried a

formaldehyde modification resulting in a mass increase of 30Da [Metz et al., 2004; Metz et al., 2006; Toews et al., 2008].

To assess if the native SEC and crosslinking SEC approaches were biased towards the detection of proteins in specific subcellular compartments, the proteins identified in the two datasets were compared to a published dataset [Larance et al., 2013]. In the reference dataset, cells were subjected to subcellular fractionation prior to mass spectrometry based protein identification in each separate fraction [Larance et al., 2013]. The degree of overlap between the proteins identified in my datasets and the reference dataset, for each subcellular compartment, are presented as a percentage in Figure 4.8A. This demonstrates that the crosslinking SEC approach led to the detection of more cytoskeletal, membrane and nuclear proteins than the native PBS-based approach. Although the PBS approach was slightly biased towards the detection of cytoplasmic proteins, the overall proportion of proteins detected across the subcellular fractions between the two SEC approaches closely matches the distribution of proteins in the reference dataset (Figure 4.8B).

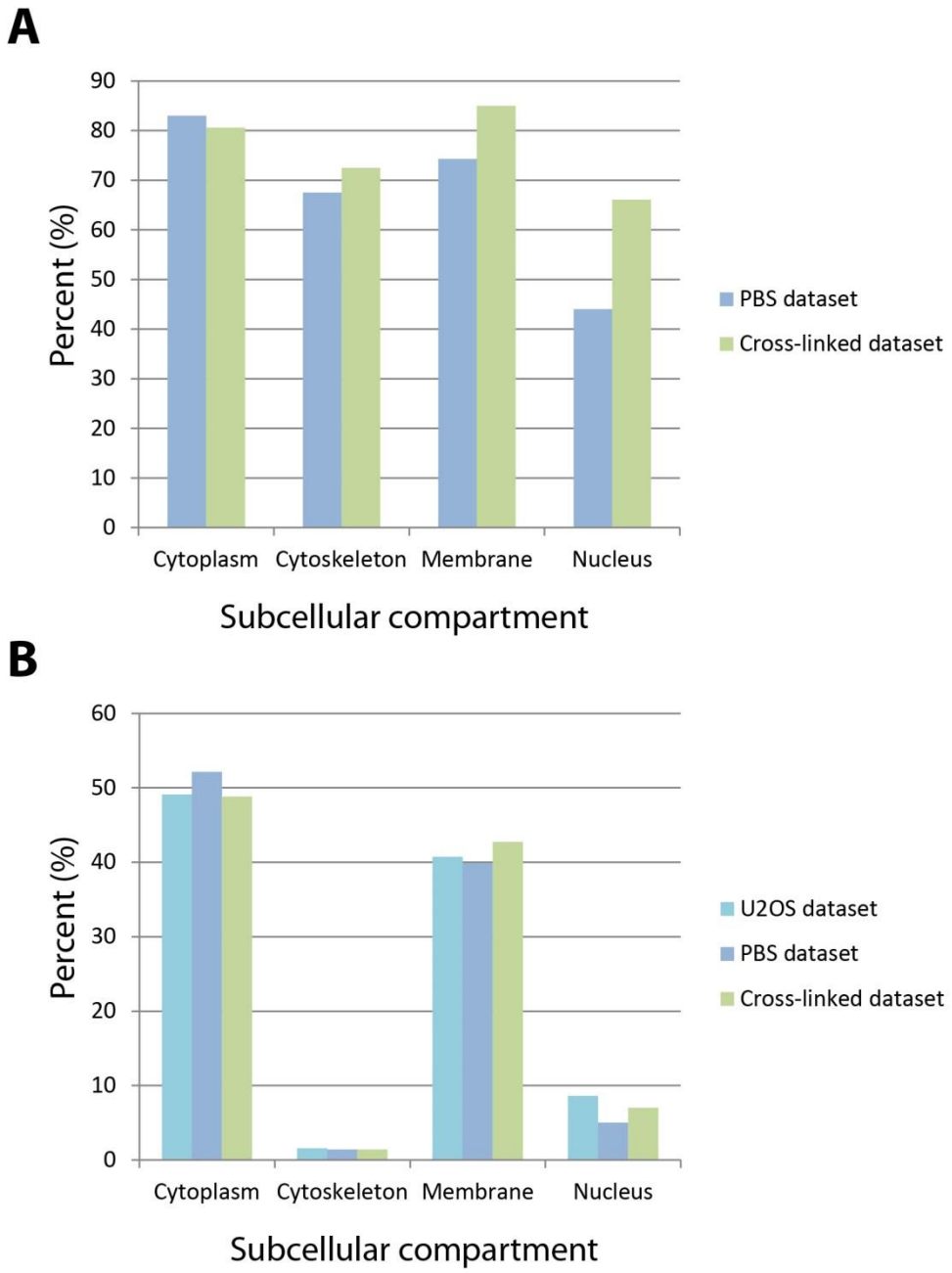


Figure 4.8. Analysis of identified proteins across different subcellular compartments.

In the reference dataset, taken from Larance *et al.* [Larance et al., 2013], a protein was allocated to the subcellular compartment where 80% or more of the protein was detected. Proteins distributed more generally across compartments, outwith these criteria, were excluded from the analysis. Proteins identified in the two SEC datasets were then compared to the remaining 2,531 proteins in the reference dataset. *A*, bar graphs demonstrating the percentage overlap of proteins identified in each of the SEC datasets, compared to the reference dataset, for each subcellular fraction. *B*, bar graphs demonstrating the proportion of proteins allocated to each subcellular compartment as a percentage of the matched proteins for the SEC datasets and the reference dataset.

4.2.5 The crosslinked SEC approach is reproducible

To determine the reproducibility of protein complex separation using formaldehyde crosslinking and SDS based cell lysis with size exclusion chromatography, a pairwise comparison of the proteomic analysis of all the proteins in each dataset across each of the three biological replicates was performed (i.e. replicate 1 vs 2, 1 vs 3 and 2 vs 3). In each comparison a Pearson correlation coefficient was calculated, to compare each normalised protein elution profile across the 48 SEC fractions. Each comparison is shown in a density plot that reveals a high correlation (>0.7) between replicates for the majority of proteins (Figure 4.9).

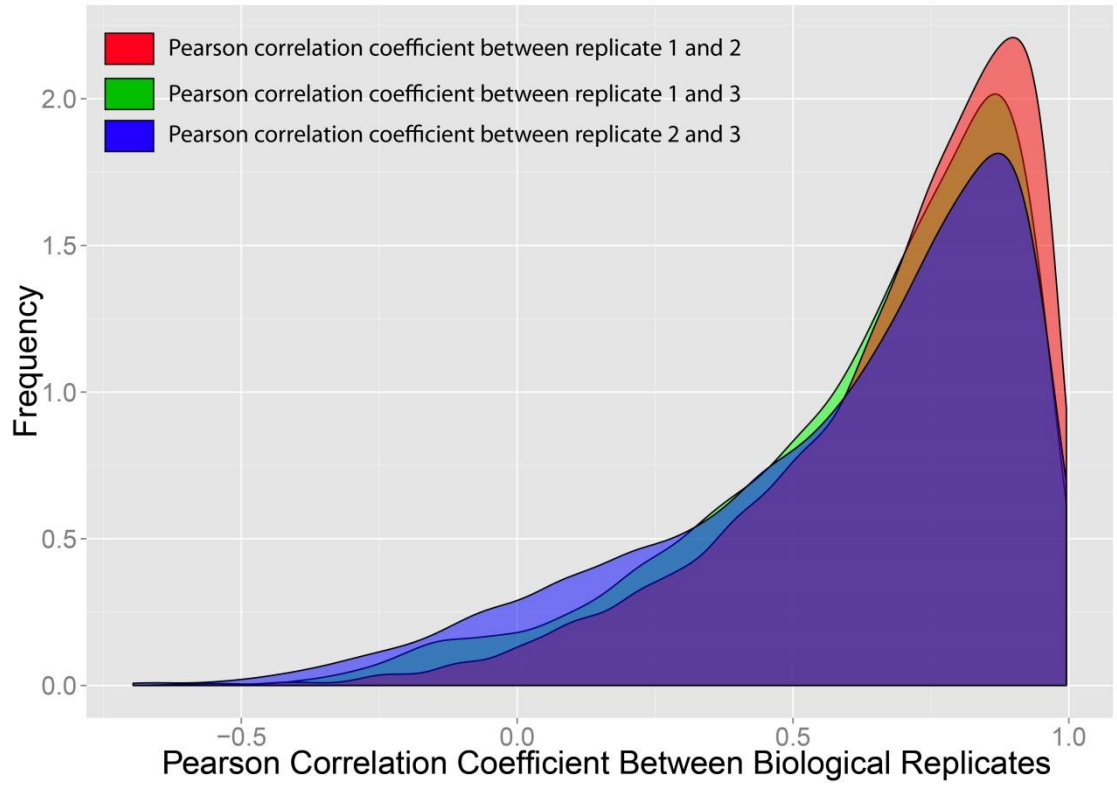


Figure 4.9. Reproducibility of the crosslinked SEC elution profiles.

The Pearson correlation coefficient was calculated for the three biological replicates (replicate 1 *versus* replicate 2, replicate 1 *versus* replicate 3, and replicate 2 *versus* replicate 3). The relative protein density (y-axis) was plotted against the Pearson correlation coefficient for each replicate (x-axis).

4.2.6 Hierarchical clustering of protein elution profiles

As before, proteins interacting within the same complexes were expected to have similar elution profiles across SEC fractions. Therefore, hierarchical clustering was performed to identify potentially interacting proteins. Truly interacting proteins would be expected to be crosslinked and therefore co-elute in the same pattern in each of the three biological replicates. As a result, the clustering analysis was performed across the 144 fractions rather than the mean of the three replicates. The generation of 1 to 1,000 clusters using the crosslinked SEC dataset was performed to determine the optimum number of clusters required. The correlation co-efficient for proteins within each cluster, based on the similarity of elution profiles, was calculated to determine the optimum cluster size (Figure 4.10). Forming 750 clusters kept the similarity of elution profiles within a cluster high (correlation coefficient >0.85), whilst minimising the subdivision of known interacting proteins. The heat map of the 750 clusters across three replicates is presented in Figure 4.11.

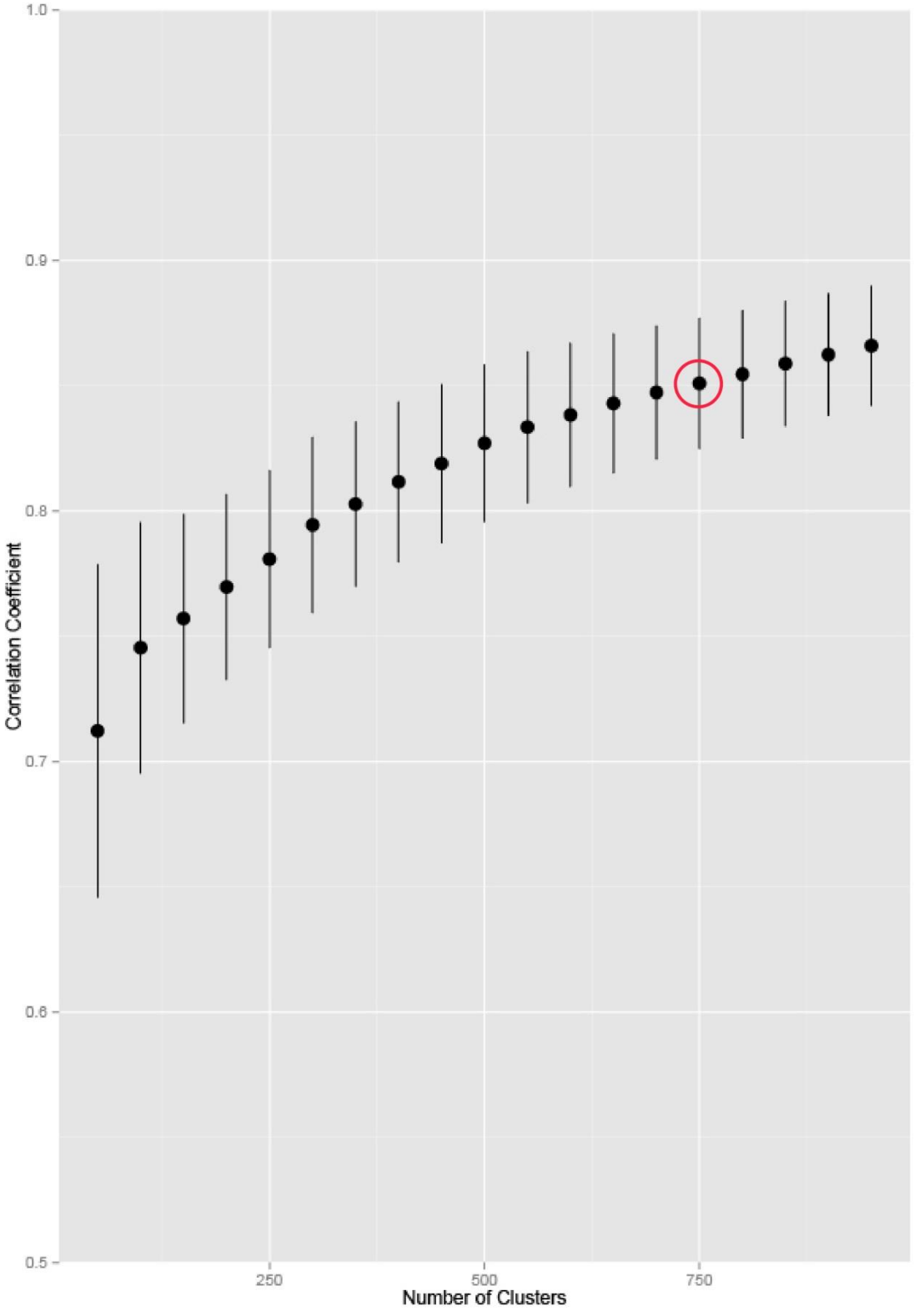


Figure 4.10. Calculation of the optimum hierarchical cluster number.

The column graph shows the average Pearson correlation co-efficient (y-axis) plotted over a range of cluster numbers between 1 and 1000 (x-axis). The red circle highlights the correlation coefficient for 750 clusters.

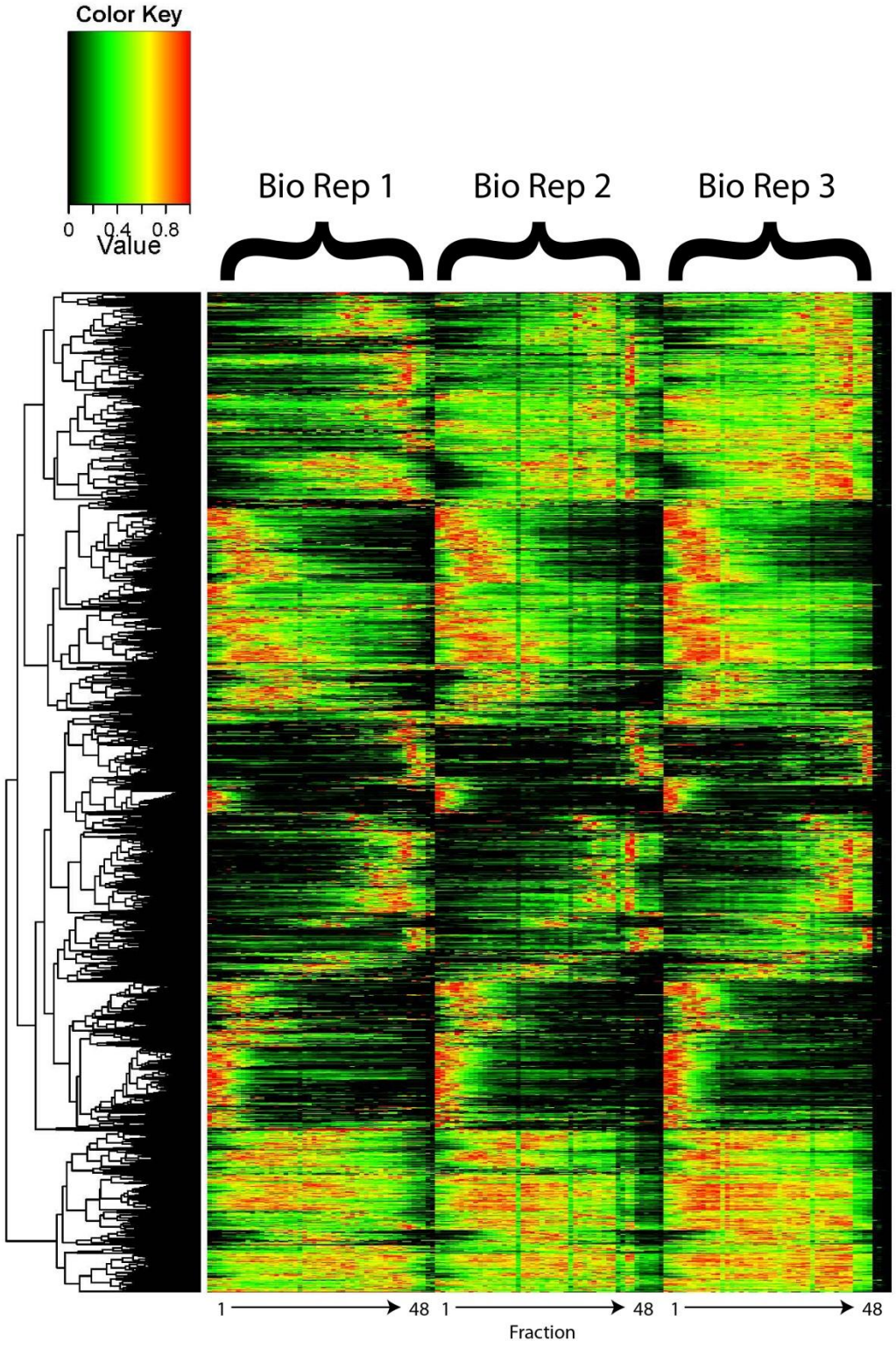


Figure 4.11. Hierarchical clustering of protein elution profiles.

All 6,425 proteins were clustered on the basis of the similarity of their elution profiles across 144 SEC fractions and the results are presented as a heat map. The dendogram cut-offs for 750 clusters are illustrated to the left of the heat map. The key shows the colours allocated to the normalised spectral count as it varies from 0 (black) to 1 (red).

4.2.7 Components of known protein complexes co-elute within the same cluster

Using the STRING database to analyse proteins clustered together, components of known protein complexes can be identified. For example, closer analysis of cluster 167 shows it to contain multiple components of the minichromosome maintenance (MCM) complex (MCM2, MCM3, MCM4, MCM5 and MCM7) (Figure 4.12). This complex unwinds duplex DNA and powers replication fork progression during DNA replication [Bochman and Schwacha, 2009]. This cluster has an overall peptide elution peak between fractions 8-10, corresponding to an approximate size of >1.8MDa, although it is noted that the MCM proteins strongest elution peak is in earlier fractions compared to other proteins in this cluster. A previous study investigating the archaeal MCM complex by SEC has demonstrated the elution peak to be in a region corresponding to a dodecamer (~950kDa) [Pape et al., 2003]. Similarly, the MCM2-7 complex has been shown to form a double hexamer in xenopus extracts, with an elution peak seen in the region of 1MDa [Gambus et al., 2011]. The MCM complex has also been shown to interact with Cdc45 and GINS to form the CMG complex [Gambus et al., 2011]. Interestingly, both Cdc45 and GINS complex proteins both have elution peaks in fractions 1-4. Although the overall elution profiles of these interacting proteins are different, it is possible that the CMG complex is crosslinked in my dataset, explaining why the MCM complex eluted earlier than would have been expected for a double hexamer of this complex alone. It is also likely that DNA fragments remained intact through cell lysis and sonication therefore the elution of this complex in a fraction with such a large predicted molecular weight, may represent it being bound to residual DNA.

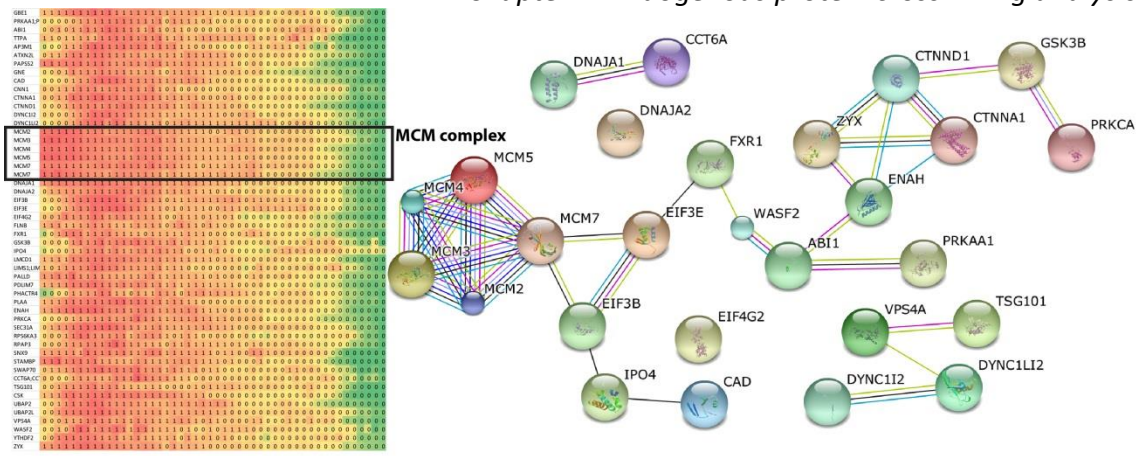


Figure 4.12. Components of the MCM complex co-elute.

The heat map for cluster 167 is presented and the profiles for the MCM complex proteins are highlighted in the black box (left panel). The normalised spectral counts are represented by green colouring for 0 and red for 1. The string diagram for the co-clustering proteins demonstrates known protein-protein interactions.

Analysis of cluster 97 shows it contains multiple subunits of the CCT complex (CCT1/TCP1, CCT2, CCT4, CCT5 and CCT7) (Figure 4.13). Crosslinking structural analysis of this complex shows the complex comprises two stacked octameric rings, each containing CCT subunits numbered 1-8. Interestingly, CCT1,4,2,5,and 7, the proteins identified within cluster 97, all interact as adjacent subunits [Kalisman et al., 2012].

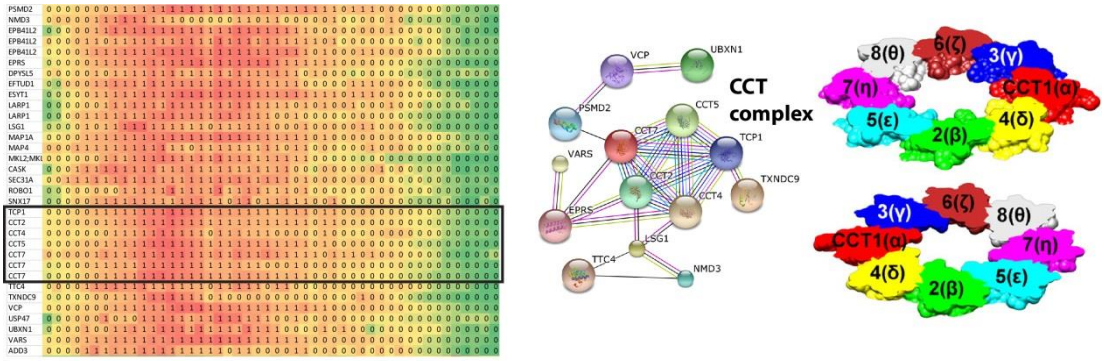


Figure 4.13. Neighbouring subunits can be crosslinked.

The heat map for cluster 97 is presented and the profiles for components of the CCT complex are highlighted in the black box (left panel). The normalised spectral counts are represented by green colouring for 0 and red for 1. The string diagram demonstrates that known interacting proteins co-cluster (middle panel). Structure of the octameric rings of the CCT complex [Kalisman et al., 2012].

A third example cluster is number 130 (Figure 4.14). This cluster contains two out of three components of the mTORC2 complex, namely mTOR and RICTOR which are involved in cytoskeleton regulation [Dos et al., 2004]. In the same complex, TSC1 and TSC2 are also present. These proteins interact and act as negative regulators of the mTOR pathway [LaSarge and Danzer, 2014]. A second set of interacting proteins in this cluster includes IKBKG (NEMO), TRAF6, RBCK1 and ERC1. TRAF6 is involved in ubiquitination of NEMO, a regulatory subunit that inhibits NF-κB activation [Bonizzi and Karin, 2004; Zotti et al., 2014]. RBCK1 is an E3 ubiquitin ligase in the linear ubiquitin chain assembly complex (LUBAC), a complex shown to conjugate linear polyubiquitin chains to NEMO to regulate NK-κB activation [Tokunaga et al., 2009]. ERC1 is a regulatory subunit of IκB kinase, known to interact with NEMO [Sigala et al., 2004]. As demonstrated, all of these proteins are involved in regulation of the NF-κB pathway, but to date have not been unified as part of one complex. These examples also demonstrate that different sets of interacting proteins can elute within the same cluster.

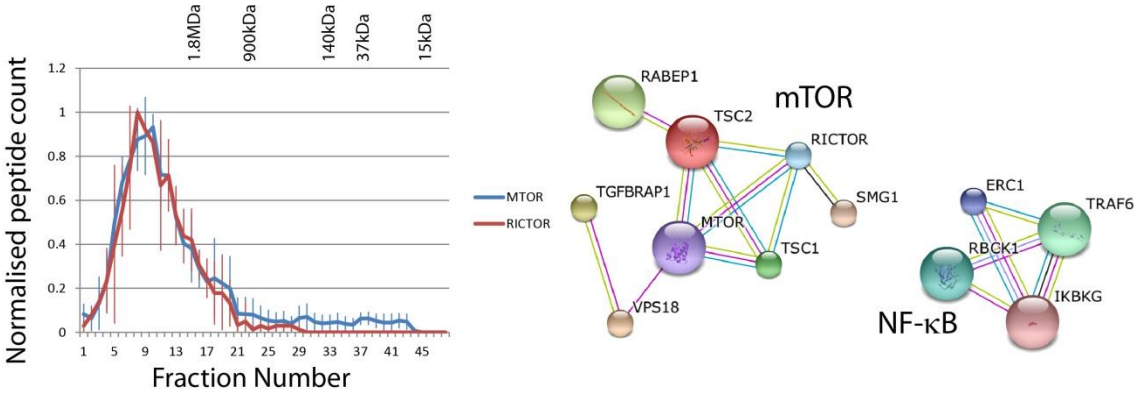


Figure 4.14. Independent protein complexes can co-cluster.

Line graph of mTOR and RICTOR with normalised peptide count (y-axis) plotted for each fraction (x-axis) (left panel). Error bars indicate the standard error from the mean. The STRING interaction networks for the mTOR and NF-κB co-clustering proteins are shown in the right panel.

Analysis of these clusters demonstrates that protein crosslinking with formaldehyde can hold interacting proteins together in complexes through SDS denaturation and gel filtration.

4.2.8 The crosslinking SEC approach enhances the identification of transmembrane proteins

One of the key aims in developing the crosslinking strategy was to improve the detection of factors such as membrane proteins and their complexes, which were under-represented using the native extraction conditions, by using a highly denaturing lysis buffer. As shown in Figure 4.8, the overall detection of membrane proteins in this dataset was higher than the PBS approach. Proteins detected in the crosslinked dataset that were not detected in the native complex analysis included numerous syntaxin proteins and other components of the SNARE (soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor) complex. SNARE proteins assemble into membrane-bridging complexes [Fasshauer, 2003] and are involved in intracellular trafficking [Ungermann and Langosch, 2005]. Other transmembrane proteins such as GLUT1, a glucose transporter with 12 transmembrane segments were also identified and resolved using the crosslinking combined with denaturing cell lysis technique [Deng et al., 2014]. This type of protein would be difficult to analyse using a native buffer. Another example of such a complex is represented by integrins. In the crosslinked dataset, many integrins co-clustered in cluster 19 (Figure 4.15). These are transmembrane proteins, the majority of which were not identified in the non-crosslinked SEC dataset [Campbell and Humphries, 2011]. Integrins play a vital role in sensing changes in the extracellular environment, which enables cells to adapt and

modulate proliferation, differentiation and survival [Wolfenson et al., 2013]. Integrins form dimers between alpha and beta subunits [Campbell and Humphries, 2011], alpha subunits ranging in size from 130-210k and beta subunits 95-130kDa [Hynes, 1987]. This cluster contains many alpha subunit proteins (alpha-2, 3, 5, 6, 7, V) and one beta subunit protein (beta-3). The elution peak for this cluster is ~500kDa, possibly representing the size of an intact heterodimer with the additional size contributed by glycosylation of each subunit [Wolfenson et al., 2013], or co-elution with one of the many proteins known to interact with integrins such as Src [Shattil, 2005], CD98hc [Fenczik et al., 2001] and Talin 1 [Ye et al., 2014].

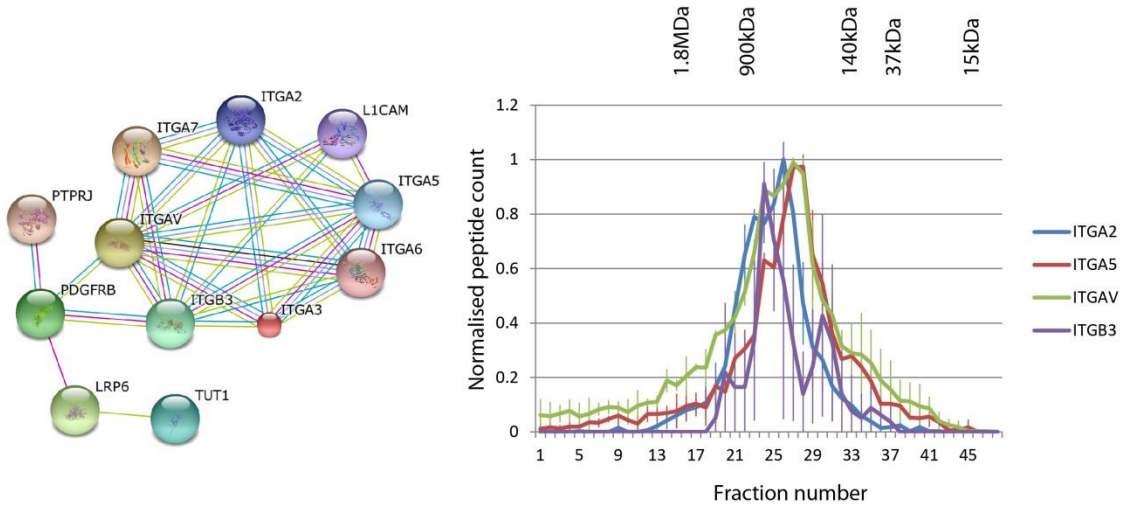


Figure 4.15. Transmembrane proteins are identified and co-cluster.

STRING interaction network of co-clustering proteins, including many integrins, in cluster 19 (left panel). Line graphs for representative alpha (ITGA2, ITGA5, ITGAV) and beta (ITGB3) subunits contained within cluster 19 (right panel). The normalised peptide count (y-axis) is plotted for each fraction (x-axis). Error bars represent the standard deviation from the mean.

4.2.9 Crosslinking with formaldehyde does not crosslink all subunits of the proteasome

On reviewing the clustering data, it was apparent that some proteasome subunits clustered differently. Alpha subunits and regulatory subunits demonstrate overlapping elution profile peaks around fraction 15, whereas beta subunits are only weakly identified here (Figure 4.16). The main elution profile peak for beta subunits forms much later, from fraction 25 onwards. Interestingly, some alpha subunits and regulatory cap components also peak in these later fractions, but the approximate size of complexes eluting in these fractions is too small for the intact proteasome (see Figure 4.16 for representative examples).

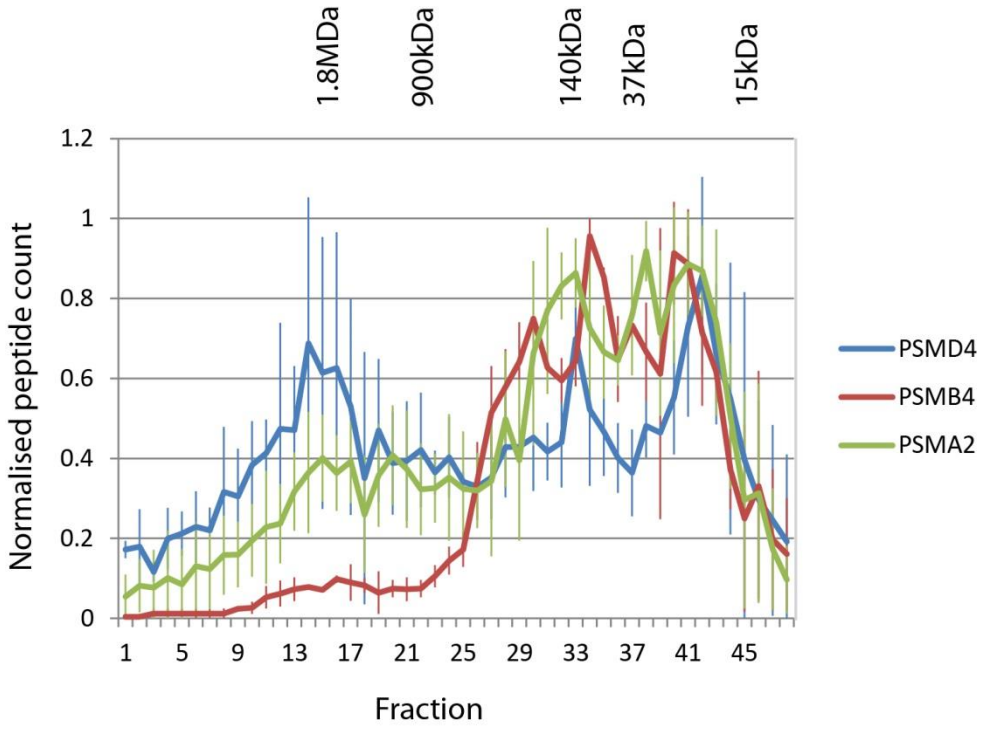


Figure 4.16. Analysis of proteasome subunits.

Line graph demonstrating the normalised peptide count (y-axis) for each fraction (x-axis) for selected alpha (PSMA2), beta (PSMB4) and regulatory cap (PSMD4) proteasomal subunits. Error bars represent the standard deviation from the mean.

Analysis of proteasome structure from the literature shows that the 20S core particle comprises two outer rings of alpha subunits and two inner rings of beta subunits [Kunjappu and Hochstrasser, 2014]. The 19S regulatory cap has a base that sits on the alpha ring and a lid attached to the cap [Sharon et al., 2006] (Figure 4.17). The SEC profile peaks seen around fraction 15 above, correlate closely with the structural data and may represent an intact complex of the alpha subunit and regulatory cap. Partially crosslinked components comprising regulatory cap, alpha and beta subunits eluted later from the column.

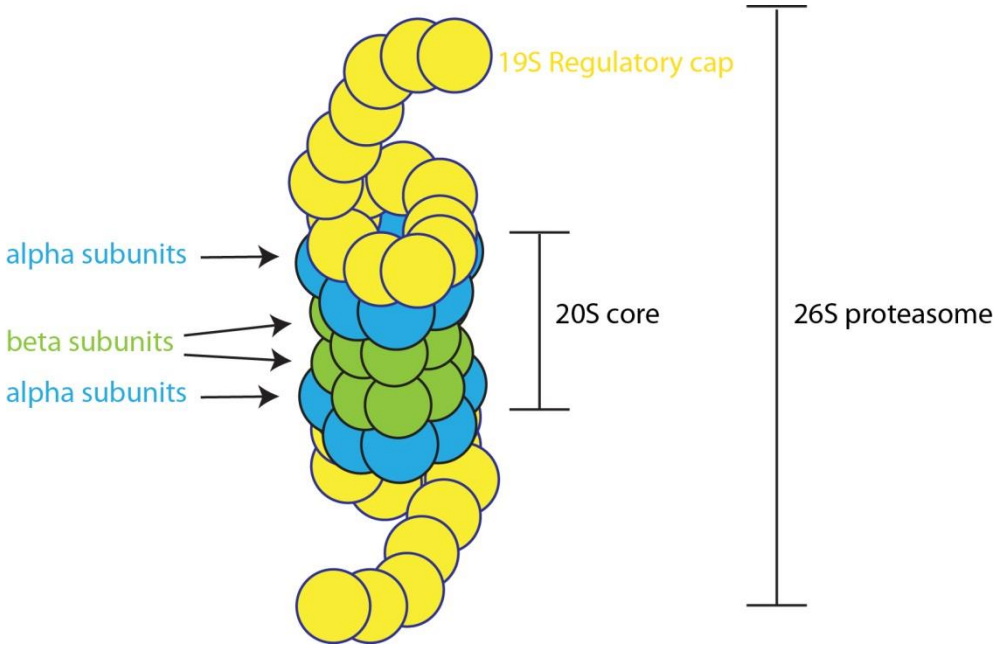


Figure 4.17. Structure of the 26S proteasome.

Schematic diagram of the 26S proteasome, formed by the 20S core and the 19S regulatory cap.

4.3 Discussion

In this chapter I have generated a comprehensive dataset of the protein interactions in U2OS cells by protein crosslinking, denaturing cell lysis and size exclusion chromatography. This method has proved very effective for the analysis of a wide

variety of protein complexes and has improved the detection of membrane proteins over the PBS based approach.

In this study, I elected to use formaldehyde as the crosslinker but as discussed previously (section 4.1.1) there are alternative crosslinking agents that could be utilised. Formaldehyde crosslinks are less specific in their targets than other crosslinkers, which was deemed to be an advantage in this study as it would enable a wider range of proteins to be crosslinked under the same conditions. A disadvantage, however, is that the spacer arm is very short (2.3-2.7Å) and therefore, interacting proteins need to be in very close contact for them to be crosslinked and contain residues that react with formaldehyde. This problem possibly contributed to the partial crosslinking of the proteasome beta subunits to the alpha subunit and regulatory cap. Either increasing the concentration of formaldehyde, or increasing the length of crosslinker incubation time, may never improve the crosslinking of complexes if their structure means reactive sites do not lie within the size of the spacer arm of the crosslinking agent. In future experiments, performing the analysis using a different crosslinker, such as DSP, which has a 12Å spacer arm, would perhaps improve the detection of intact, large complexes.

I elected to use 6% formaldehyde concentration to crosslink proteins *in vivo* in this study as provisional experiments demonstrated this held known multimeric proteins in complexes. This concentration is higher than the 1% used in ChIP [Kuo and Allis, 1999] and immunoprecipitation studies [Vasilescu et al., 2004; Guerrero et al., 2006; Tagwerker et al., 2006]. However, I have demonstrated that for some multimeric proteins, 6% formaldehyde concentration was required to sufficiently retain protein crosslinks under denaturing conditions. It is possible that this concentration over

crosslinked some proteins into non-specific complexes. However, these non-specific crosslinks would be expected to be more random than specific crosslinks and would not be reproduced across the triplicate analysis. None the less, additional analysis of crosslinking protein complexes in the presence of varying formaldehyde concentrations over different timescales could be beneficial in the future and would add confidence in interactions if proteins continued to co-elute together.

One of the key aims in developing the crosslinking strategy was to improve both the overall efficiency of protein isolation in cell extracts and more specifically to enhance the detection of factors, such as membrane proteins and their complexes that are under-represented in extracts prepared using native extraction conditions. The extracts from crosslinked cells typically have higher total protein concentration than native extracts, consistent with a higher efficiency of protein isolation resulting from the solubilising effects of detergent. Although Figure 4.8 indicates that most membrane proteins were also detected in the native extract dataset, I propose that many of these membrane proteins were not well resolved by SEC when extracts were prepared using the PBS buffer. For example, GLUT1 is a glucose transporter with 12 transmembrane domains [Deng et al., 2014]. It is known to form a complex with ADD2 and DMTN through which it is linked to the actin cytoskeleton [Khan et al., 2008]. On the basis of molecular weights, this complex would be predicted to have an approximate molecular weight of ~177kDa. In the crosslinked dataset, the GLUT1 elution profile has a peak around fraction 31, which would correlate with it participating in a complex of just above 140kDa. GLUT1 was also identified in the PBS dataset, but here its main elution peak was in a fraction corresponding to an approximate molecular weight of ~1MDa. In a native buffer, with no detergent

solubilisation, proteins closely associated with the cell membrane are unlikely to be efficiently extracted. However, as a result of sonication, these proteins may be contained within lipid vesicles. Consequently, during native complex analysis, these vesicle-associated proteins are likely to be predominantly in or near the void fraction and hence not efficiently resolved as separable complexes. I propose therefore that the use of in vivo crosslinking and denaturing cell lysis has not only improved the detection of membrane proteins, but has also enhanced the ability to resolve the cognate protein complexes using SEC fractionation.

As discussed by Salwinski and Eisenberg, to fully utilise the wealth of information obtained from interaction networks, models need to incorporate data regarding the spatial and temporal dynamics of proteins, as well as PTMs and degradation pathways [Salwinski and Eisenberg, 2003]. In the Lamond laboratory, studies have been performed to analyse differential protein expression [Ahmad et al., 2012] and degradation [Larance et al., 2013] across subcellular compartments, and differential protein expression across the cell cycle [Ly et al., 2014]. These studies were performed at the level of individual proteins but to date, such analyses have not been attempted at the level of protein complexes. The PBS-SEC dataset demonstrates how protein isoforms and PTMs can participate in different protein complexes. However, other methods of studying native protein complexes, such as native SEC, ion exchange and sucrose gradients are limited in their ability to identify truly native complexes in different subcellular compartments as the chemical buffers required to isolate each fraction are likely to disrupt PPIs. However, the crosslinking technique could easily be combined with subcellular fractionation, because the crosslinks would remain intact through cell lysis, including variable salt concentrations and denaturing conditions. This

would then enable the analysis of protein complex distribution between subcellular compartments to give a more complete picture of biological processes within the cell. SEC separation of each subcellular fraction would also be likely to increase the overall protein identifications, increasing the depth of coverage in the dataset.

Bottom-up MS approaches combined with crosslinking can be limited by missed cleavages [Sinz, 2006; Petrotchenko and Borchers, 2010]. As formaldehyde crosslinks form at the site of primary amine groups at lysine residues and at the N-termini of proteins, it is possible that trypsin (the most commonly used protease) cannot cleave C-terminal to the modified residue. This results in large peptides, which may not be identified. Additionally, the loss of a positive charge after modification of the amino group by crosslinking can result in products with low charge states being created during electrospray ionisation. These ions may have too low a charge state for mass spectrometry analysis and as a result, would not be detected [Sinz, 2006]. However, studies utilising formaldehyde in other crosslinking strategies combined with MS suggest that missed-cleavages are not problematic [Shevchenko et al., 1996; Vasilescu et al., 2004; Tagwerker et al., 2006; Sutherland et al., 2008]. In addition, reversal of crosslinks prior to protease digestion in my study will have reduced the size of the resultant peptides, thereby reducing the impact of these limitations on peptide detection by MS. Despite these potential limitations, I detected more than 88,000 peptides. To improve on this, additional proteases such as cyanogen bromide could be added to the workflow. Cyanogen bromide cleaves at the C-terminal side of methionine, a site not modified by formaldehyde. Therefore, larger peptides not cleaved with trypsin might be cleaved by this protease instead, thereby improving detection rates. Improvement in instrument technology through advances in mass

resolution and mass accuracy could overcome some of the complications raised by crosslinking and mass spectrometry. Top-down approaches for future crosslinked protein analysis is a possibility, as intact complexes could be analysed in the mass spectrometer. However, ionisation and fragmentation capabilities of the instruments can limit the size of the complex being analysed [Petrotchenko and Borchers, 2010]. Novak *et al.* have combined a top-down and bottom up approach, cleaving intact protein complexes into larger peptide fragments using cyanogen bromide, and then ionising these larger crosslinked peptide fragments for top-down analysis [Novak et al., 2005]. This approach might be worth considering in the future for a greater depth of coverage of crosslinked peptides.

In this study, hierarchical clustering was performed to identify interacting proteins based on the similarity in elution profiles. As many proteins can take part in multiple different complexes, this clustering technique can limit the analysis of the different complexes. Work is ongoing in the Lamond laboratory to develop analysis software that combines hierarchical clustering with "peak picking". In this approach, each peptide peak across an elution profile would be identified separately and matched to other proteins, which also peak in the same fraction. For example, VCAM-1/L1CAM was clustered in my dataset with several alpha-integrins and beta-3 integrin. However, VCAM-1 is known to be recognised by Leu-Asp-Val (LDV)-binding integrins such as $\alpha 4\beta 1$ [Humphries et al., 2006; Humphries et al., 2009]. Beta-1 integrin was identified in the crosslinking dataset, but in a different cluster, even though the elution profile for this protein contains a peak which corresponds to that seen in the elution profile for VCAM-1. By combining hierarchical clustering and "peak picking" techniques these proteins may have been identified as interaction partners. Therefore, by using this

technique, it is envisaged that a higher number of protein interactions could be identified.

Taken individually, this dataset and the PBS based dataset from chapter 3, provide powerful tools for the study of protein complexes, particularly when used in conjunction with affinity purification studies. However, cross-referencing either analogous studies [Byron et al., 2012], or experimental data with interaction databases [Jacquemet et al., 2013], enables the determination of whether a protein of interest is a component of one or more protein complexes. Such a cross-referencing approach also enables the identification of additional complex components and increases confidence in the likelihood of a true PPI. Therefore, with the accumulation of more datasets based on the same technique, it should be possible to extend the use of my datasets to more accurate PPI predictions. For example, if a pair of proteins clustered together in the PBS dataset and also clustered together in the crosslinked dataset, the confidence that these two proteins were truly interacting is higher than if they only co-clustered in one of the two datasets. If a scoring system could be devised to calculate a probability interaction score for all proteins identified across the two datasets, more PPIs, and therefore novel protein complexes, could be predicted with increased confidence. A similar approach was applied by Havugimana *et al.*, who included predicted interactions from evolutionary data, mRNA co-expression and domain co-occurrence in addition to their chromatographic and sucrose gradient approaches [Havugimana et al., 2012]. If such analysis could be performed using my data in the future, a comparative study of predicted interactions between our two approaches would be of great interest.

4.4 Distribution of Effort

The SEC and mass spectrometry sample runs were performed by Dr Mark Larance. The R-based statistical analysis was also performed by Dr Mark Larance. Dr Yasmeen Ahmad performed the statistical analysis of protein subcellular localisation.

Chapter 5: Analysis of the MRFAP1-MORF4L1 interaction

5.1 Background

PTMs play a major role in maintaining cell integrity and function through the regulation of protein homeostasis – as discussed in Chapter 1. As a result, pharmaceutical companies are targeting components of PTM pathways in the development of new therapeutic agents. One such compound is MLN4924, a small molecule inhibitor that selectively targets the E1 NEDD8-activating enzyme [Soucy et al., 2009]. This drug causes DNA re-replication, cell cycle arrest and resultant apoptosis. For this reason it is now in Phase 1 clinical trials for the treatment of acute myeloid leukaemia (AML) and other non-haematological malignancies [Swords et al., 2010]. However, there is still limited understanding of the detailed biochemical effects of MLN4924 on the proteome. A quantitative proteomic study utilising stable isotope labelling with amino acids in cell culture (SILAC) performed in the Lamond laboratory has demonstrated that treatment of U2OS cells with the MLN4924 inhibitor causes up-regulation of a protein called MORF4 family associated protein (MRFAP1) and one of its known interaction partners, MORtality Factor 4-like protein 1 (MORF4L1), amongst other proteins [Larance et al., 2012]. Analysis of MRFAP1 and MORF4L1 interaction partners through a series of immunoprecipitation experiments suggests that the MRFAP1-MORF4L1 interaction occurs as a mutually exclusive event to MORF4L1-MRGBP binding.

5.1.1 MRFAP1

MRFAP1 is a 14kDa protein containing 127 amino acids. Analysis of its domain structure shows it to have an N-terminal MRG-binding domain, a glycine rich flexible linker region and a coiled-coil domain towards its C-terminus [Thompson et al., 1994; Kelley and Sternberg, 2009; Waterhouse et al., 2009]. MRFAP1 has not been widely studied although its interaction with MORF4L1, via the MRG domain, has been characterised in *in vitro* [Zhang et al., 2006b] and affinity purification studies [Leung et al., 2001; Pardo et al., 2002; Hayakawa et al., 2007]. It has been suggested that MRFAP1 is a component of the MORF4L1-Rb complex to activate the *B-myb* promoter [Leung et al., 2001], but confirmation of a role for MRFAP1 in this process remains to be confirmed. A later study attempted to elucidate the *in vivo* function of MRFAP1 by creating a MRFAP1-deficient mouse model, but no definitive phenotype was identified with knockout mice reamining healthy and fertile [Tominaga et al., 2004]. Interestingly, MRFAP1 and MORF4L1 were both identified as rapidly turned over proteins in recent pulse-SILAC studies [Schwanhausser et al., 2011; Boisvert et al., 2012].

5.1.2 MORF4L1

MORF4L1 is a highly conserved member of the MORtality Factor on chromosome 4 (MORF4) gene family and is expressed in a wide range of tissues. MORF4L1 is a 41kDa protein containing 362 amino acids. It has a N-terminal chromodomain that binds to di- or tri- methylated histone H3 on K36 [Zhang et al., 2006a] (Figure 5.1). Chromodomains are histone methylated-lysine recognition modules found in a relatively small number of proteins [Yap and Zhou, 2011]. In general, these domains are involved in chromatin remodelling [Yap and Zhou, 2011]. In MORF4L1, this chromodomain has been shown to be important for the formation and activity of

histone acetyltransferase and transcriptional regulation through the B-myb promoter [Pardo et al., 2002] (Figure 5.1). Evidence suggests that careful control of histone acetylation through methylated histone tails prevents inappropriate transcription within protein coding regions [Lee and Shilatifard, 2007].

MORF4L1 also has a C-terminal MRG domain, through which it interacts with many proteins including MRFAP1 [Leung et al., 2001]; MRGBP [Cai et al., 2003; Bowman et al., 2006]; Pf1 [Kumar et al., 2011], Rb [Leung et al., 2001] and mSin3A [Kumar et al., 2011] (Figures 5.1 and 5.2). It is believed that MORF4L1 binds chromatin via its chromodomain to bring proteins interacting with its MRG domain into contact with chromatin and histones.

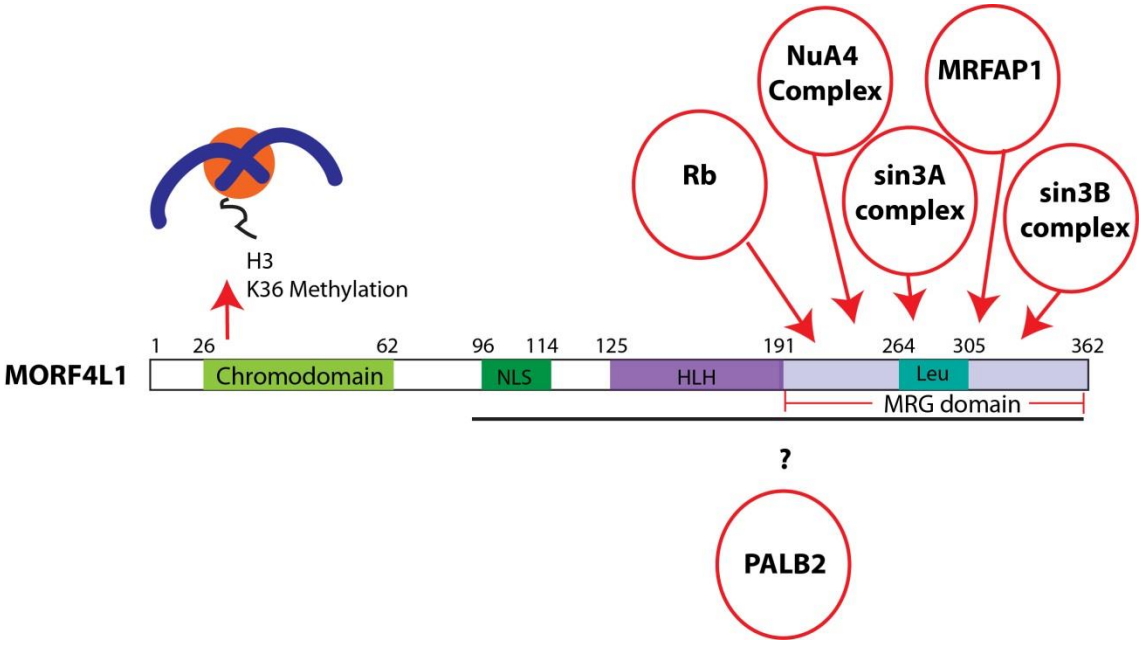


Figure 5.1. Structure of the MORF4L1 protein.

Red arrows indicate where the proteins and protein complexes, in red circles, interact with MORF4L1.

MORF4L1 has roles in embryonic development and cell proliferation [Tominaga et al., 2005; Chen et al., 2009]. MORF4L1 has been shown to be involved in activation of transcription as a subunit of the histone acetylation NuA4/Tip60 complex through its interaction with MRGBP [Cai et al., 2003]. It has also been demonstrated to play a role in repression of transcription, as part of the Rb-associated MAF1 complex [Leung et al., 2001; Pardo et al., 2002], and the histone deacetylation Rpd3S/Sin3S complex [Xie et al., 2012]. In addition, roles for MORF4L1 have also been described in homology-directed repair of chromosomal breaks through an interaction with PALB2 [Sy et al., 2009; Hayakawa et al., 2010] and in alternative splicing [Luco et al., 2010]. MORF4L1 has also been proposed to have a role in de-methylation of H3K4 through its interaction with RBP2 [Hayakawa et al., 2007]

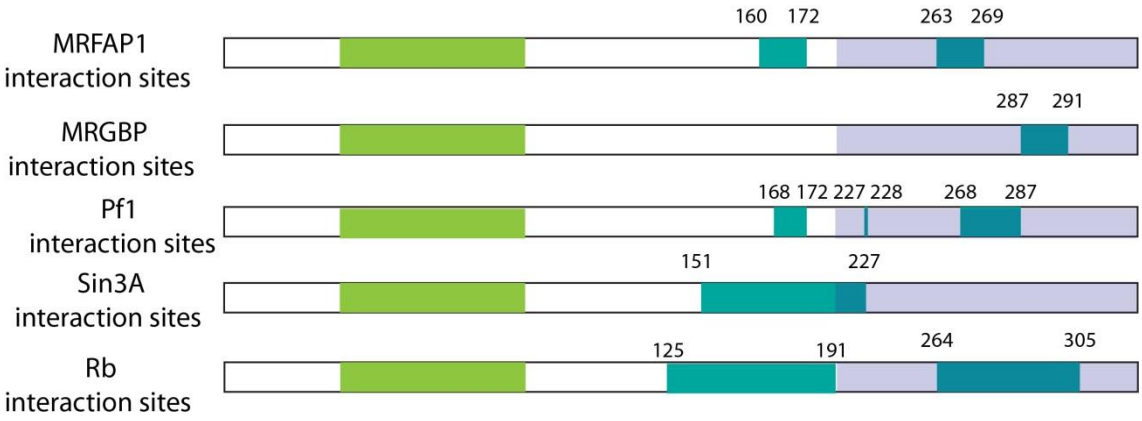


Figure 5.2. Schematic diagram indicating the interaction sites of MORF4L1 binding proteins.

The green box indicates the chromodomain. The interaction sites are indicated in turquoise. The MRG domain is represented by the grey box.

5.1.3 MORF4L1 and the NuA4 complex

The NuA4 complex is highly conserved in eukaryotes and its functions include regulating transcription, DNA repair and chromosome segregation by acetylating nucleosomal histones H4 and H2 [Doyon and Cote, 2004]. A complex comprising ~16 proteins has been characterised, including MORF4L1, MRG-domain binding protein (MRGBP), Bromodomain-containing protein 8 (Brd8) and hDomino (p400) [Doyon and Cote, 2004] (Figure 5.3).

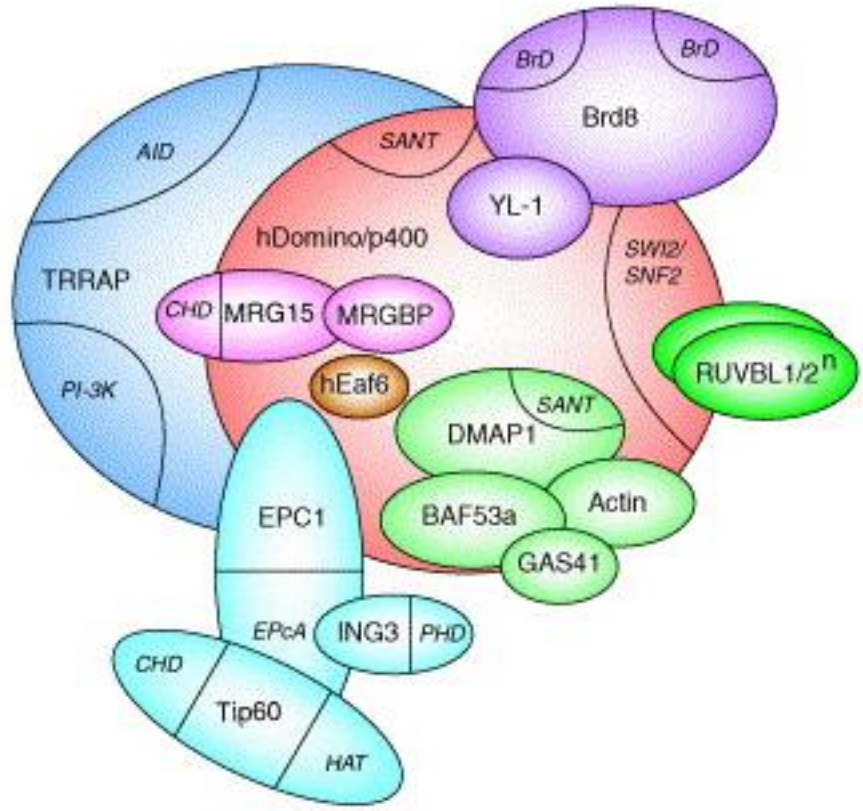


Figure 5.3. Model of the NuA4 HAT complex.

(image from Doyon and Côté [Doyon and Cote, 2004])

It has been proposed that the interaction of MRGBP and MORF4L1 is important for bringing HAT complexes into contact with chromatin [Doyon and Cote, 2004].

hDomino and Brd8 are two proteins that may function together in order to remodel chromatin through the incorporation of the variant histone H2Az into chromatin in place of histone H2A [Doyon and Cote, 2004].

Chromatin remodelling is a process that involves the disruption of histone-DNA contacts in order to expose DNA and facilitate either gene transcription, or DNA repair. Spermatogenesis in many animals is a process in which marked chromatin remodelling and histone removal occurs. This involves somatic histones being replaced by histone variants (some of which are testis specific), which are then displaced after a process of hyperacetylation, by transitional proteins. The transitional proteins are replaced by protamines in order to package DNA tightly in the sperm for successful reproduction [Govin et al., 2004; Gaucher et al., 2010]. The processes of spermatogenesis and associated chromatin remodelling are depicted in Figure 5.4.

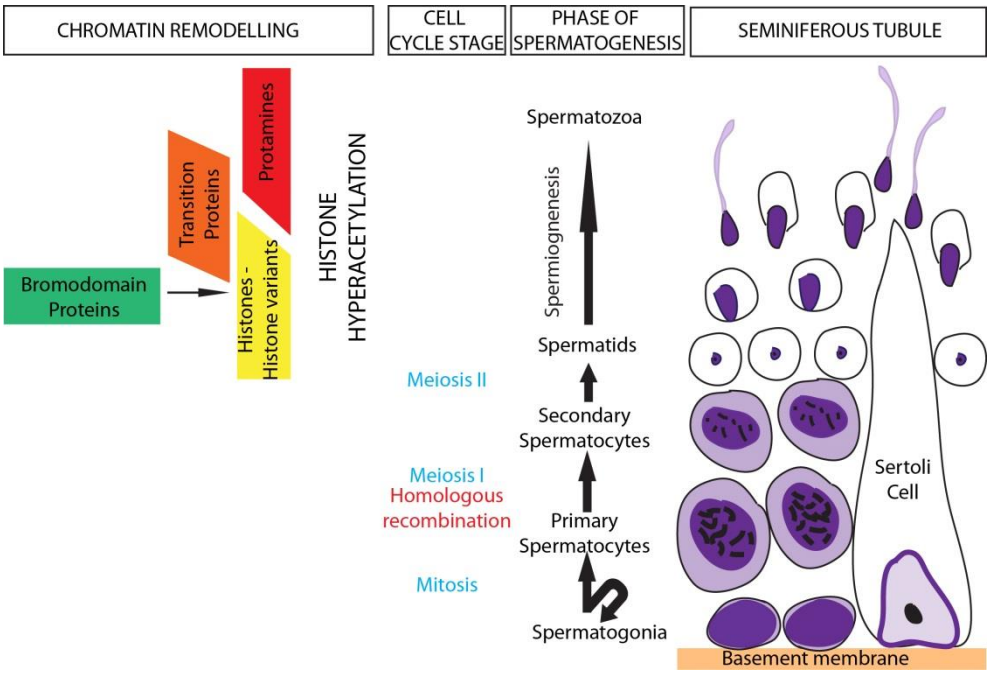


Figure 5.4. Seminiferous tubules of the testis schematic.

Illustration of the cell types present in the seminiferous tubules of the testis, the cell cycle stages they represent and how chromatin modification/remodelling is mediated during spermatogenesis.

5.1.4 Hypothesis and aims

Hypothesis

- MRFAP1 and MORF4L1 interact as a separate complex to the NuA4 complex
- MRFAP1 and MRGBP competitively interact with MORF4L1

Aims:

- To validate the SILAC screen findings for MORF4L1 and MRFAP1 using analogous techniques
- To confirm MRFAP1 and MORF4L1 form a distinct complex from the NuA4 complex utilising the native protein complex SEC technique
- To use an affinity purification strategy in combination with a SILAC screen to assess for a competitive interaction between MRFAP1 and MRGBP for the MRG domain of MORF4L1
- To analyse the immunohistochemical expression profiles of the MRFAP1 protein and its interaction partner in human tissues

5.2 Results

5.2.1 MRFAP1 is up-regulated by NEDD8 inhibition.

Prior to proceeding with any experiments to validate the MS data, I wanted to ensure I was using an antibody specific to the MRFAP1 protein, as this protein has not been widely studied by immunoblotting in the literature. To do this, siRNA knockdown of the MRFAP1 protein was performed in HaCaT (keratinocytes), U138MG (glioblastoma) and Tera-1 (pluripotent testicular carcinoma) human cell lines followed by immunoblotting (Figure 5.5).

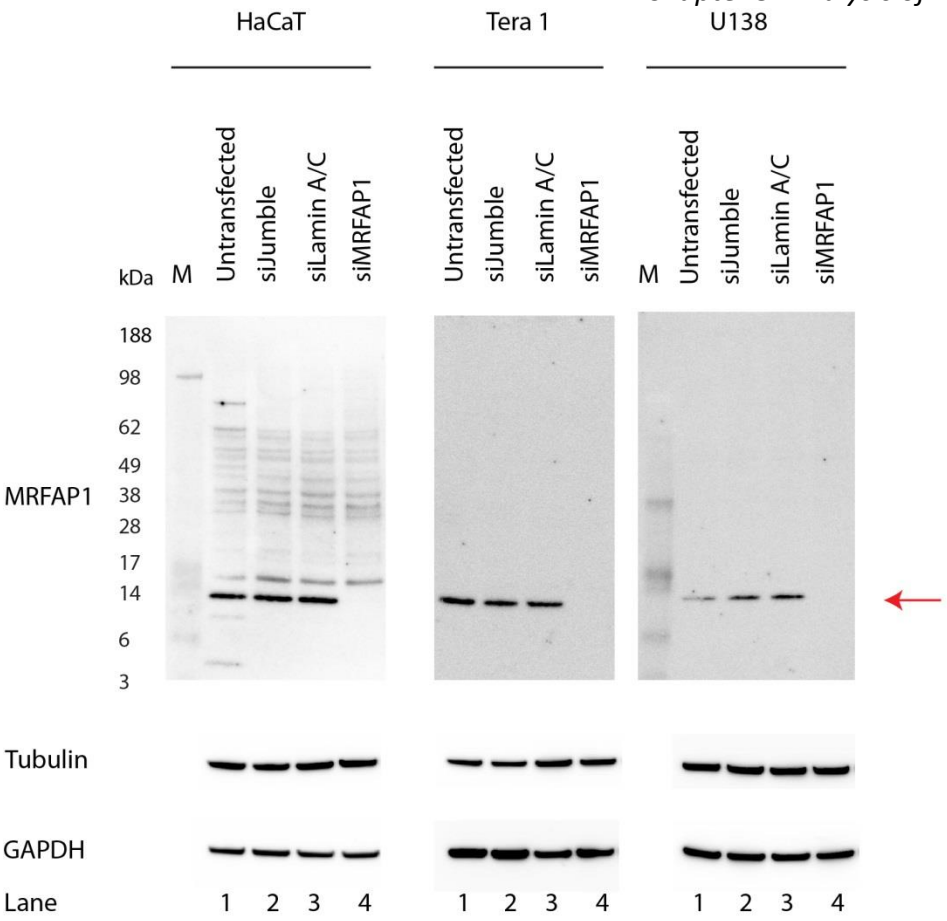


Figure 5.5. The MRFAP1 antibody shows specific detection via immunoblotting.

Immunoblotting of total cell lysates from the indicated cell types treated with the indicated siRNA for 48 hours (n=2). 10μg total protein loaded per lane. Red arrow indicates bands representing the MRFAP1 protein.

To confirm the MS data and to see if up-regulation of MRFAP1 and MORF4L1 following inhibition of NEDDylation was conserved across different cell lines, the same three cell lines were treated with MLN4924. Immunoblotting total cell lysates from each of these cell lines using antibodies specific for MRFAP1 and MORF4L1, confirmed up-regulation of these proteins in each case (Figure 5.6). On the basis of the immunoblots, MRFAP1 up-regulation is greater than MORF4L1 up-regulation in response to MLN4924 treatment, which corroborates the published SILAC data (Log_2 2.56 MRFAP1 increase and MORF4L1 Log_2 1.60).

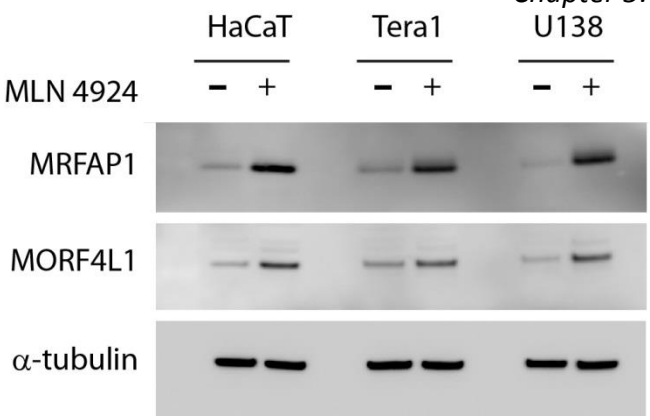


Figure 5.6. MRFAP1 and MORF4L1 are up-regulated in response to NEDD8 inhibition.

Immunoblotting of total cell lysates from the indicated cell types with 1μM MLN4924 for 18 hours or 10μl DMSO only (n=2). 10μg total protein loaded per lane.

Immunofluorescence microscopy (IF) was also performed, using HaCaT cells transfected with siRNA targeting MRFAP1 and non-coding, "jumbled" siRNA as a control, to analyse the subcellular distribution of the MRFAP1 protein. These images showed a marked reduction in the signal following MRFAP1 knockdown, demonstrating the specificity of the MRFAP1 antibody for IF (Figure 5.7). This also confirmed that in HaCaT cells, MRFAP1 had a predominantly nucleoplasmic localisation, in keeping with previous data [Leung et al., 2001]. Secondly, MRFAP1 signal intensity is vastly increased following treatment with MLN4924, consistent with up-regulation of the protein. Interestingly, cell nuclei appear to be larger following MLN4924 treatment, which is in keeping with DNA re-replication, a known effect of this compound [Soucy et al., 2009; Milhollen et al., 2011].

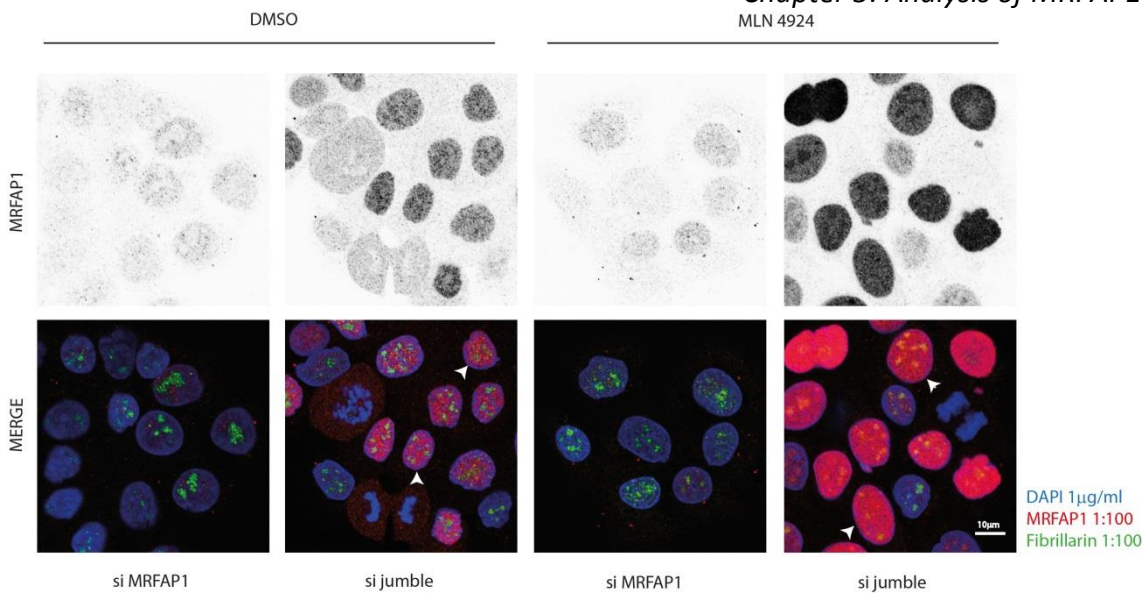


Figure 5.7. The MRFAP1 antibody shows specific detection via immunofluorescence.

Immunofluorescence microscopy of HaCaT cells treated with the indicated siRNA for 48 hours and either 1µM MLN4924, or 10µl DMSO, for 18 h (n=3). Arrowheads indicate the change in nuclear size after MLN4924 treatment. The scale bar indicates 10µm.

5.2.2 MRFAP1 is a rapidly degraded protein and stabilised by MG132

The NEDD8-Cullin E3 ligase pathway is an important regulator of protein homeostasis, with a key role in the degradation of cell cycle regulators and transcriptional control networks. As inhibition of this pathway caused up-regulation of the MRFAP1 protein, I sought to determine whether MRFAP1 is degraded via the proteasome using a cycloheximide time course to inhibit protein synthesis, combined with either MG132 (an inhibitor of the 26S proteasome), or DMSO control, in U2OS cells (Figure 5.8).

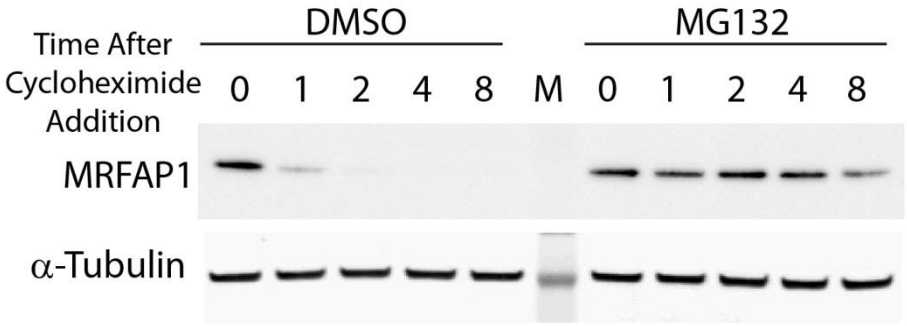


Figure 5.8. MRFAP1 is rapidly degraded via the proteasome.

Immunoblotting of total cell lysates from U2OS cells treated with 10μM MG132 or 10μl of vehicle DMSO only. A time course of 40μg/ml cycloheximide treatment was used to examine protein degradation. MG132 and cycloheximide were added at the same time for each time point (n=3). 10μg of total protein loaded per lane.

These data demonstrated that the MRFAP1 protein is degraded very rapidly, with virtually no detectable MRFAP1 protein after two hours of cycloheximide treatment. Interestingly proteasomal inhibition did not lead to a significant overall increase in MRFAP1 abundance compared to the DMSO control. However, it did show a marked decrease in the rate of MRFAP1 degradation, with significant amounts of protein still detectable after eight hours of exposure to MG132.

I noted that in the previous SILAC screen in which the MRFAP1 response to MLN4924 was reported, there were fewer proteins up-regulated by this drug than I would have expected. MG132 is known to mediate the unfolded protein response in U2OS cells, which inhibits protein synthesis by eIF2-α pSer51 phosphorylation [Jiang and Wek, 2005]. To investigate if this response was also mediated by MLN4924 treatment, thereby providing an explanation for the relatively small number of proteins up-regulated in response to this compound, U2OS cells were treated with DMSO, cycloheximide, MG132 and MLN4924 and immunoblotted to test for phosphorylation

of eIF2alpha. This confirmed that MLN4924 treatment affected cells as NEDD8 conjugation to large proteins was reduced (Figure 5.9). MG132 treatment caused upregulation of phosphorylated eIF2alpha protein as expected. However, MLN4924 did not therefore these data suggest that the UPR is not mediated by MLN4924 treatment (Figure 5.9).

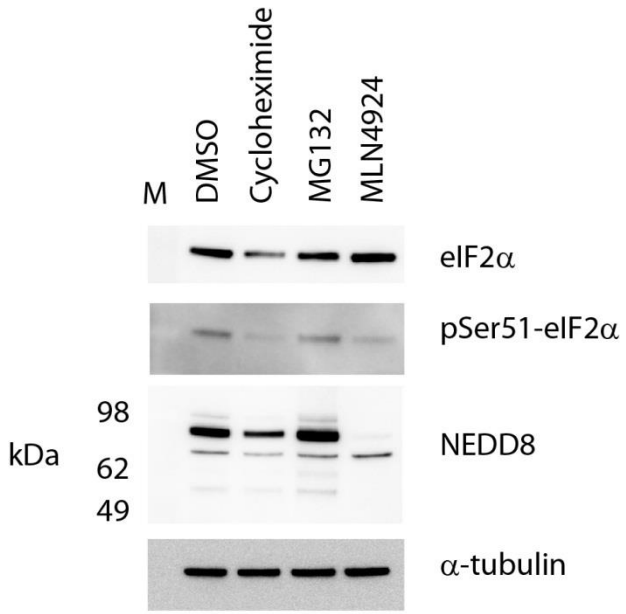


Figure 5.9. MLN4924 does not inhibit protein synthesis.

Immunoblotting of total cell lysates from U2OS cells treated with 10µl DMSO, 40µg/ml cycloheximide, 10µM MG132 or 1µM MLN4924 for 6 hours (n=3). 10µg of total protein was loaded per lane.

5.2.3 Endogenous MORF4L1 may bind MRFAP1

Although an interaction between MRFAP1 and MORF4L1 has been shown previously in *in vitro* models [Zhang et al., 2006b] and tagged affinity purifications [Leung et al., 2001; Pardo et al., 2002; Hayakawa et al., 2007], I wanted to investigate whether an endogenous interaction could be demonstrated for the first time in mammalian cells. Here, U2OS cells were treated for 16 hours with 1µM MLN4924 to increase the

endogenous MRFAP1 signal. A commercially available antibody raised against MORF4L1 was then used to immunoprecipitate the endogenous protein. Immunoblotting the immunoprecipitate did not clearly demonstrate that the endogenous MORF4L1 protein had been immunoprecipitated as the band in the IP lane was obscured by the heavy and light chains of the antibody used in the pull down. In addition, α-tubulin was co-immunoprecipitatetd with the endogenous protein which suggests the pull-down was not entirely specific. Although a clear band for MRFAP1 is seen in the IP lane for the MORF4L1 immunoprecipitation and not in the control experiment, this pull-down should be repeated with a more efficient pull down of the MORF4L1 protein, either with a higher concentration of this antibody, or, a different antibody raised against the MORF4L1 protein, to confirm an endogenous interaction between these two proteins. It may also be beneficial to use a cross-linking agent in the IP to reduce the heavy and light chain artefact (Figure 5.10).

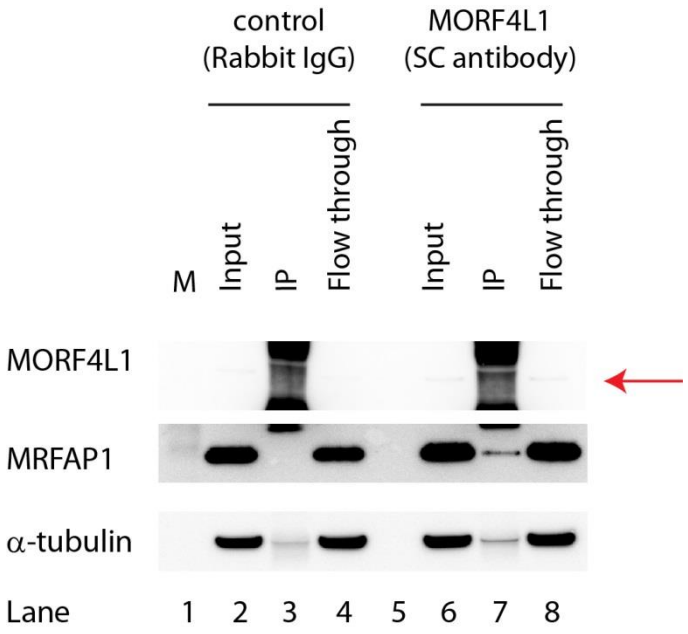


Figure 5.10. MORF4L1 may interact with MRFAP1 by endogenous immunoprecipitation.

10μg of MORF4L1 antibody was used to co-immunoprecipitate MORF4L1 interacting proteins from U2OS cells treated with 1μM MLN4924 for 16 hours

(n=2). The red arrow indicates the bands corresponding to MORF4L1. 10μg of protein were loaded per lane for the input and flow through. 10% of the immunoprecipitate was loaded in lanes 3 and 7.

5.2.4 MRFAP1 expression is stabilised by MORF4L1

As MRFAP1 and MORF4L1 are known to interact, I wanted to test whether expression of MORF4L1 affected MRFAP1 expression levels and *vice versa*. To do this MRFAP1, MORF4L1 and MRGBP endogenous proteins were knocked down using siRNA and the resultant total cell lysates were immunoblotted. This demonstrated that siRNA knockdown of MRFAP1 does not affect MORF4L1 expression (Figure 5.11A, lane 4). However, knockdown of MORF4L1 leads to a marked reduction in MRFAP1 expression, as detected by immunoblot (Figure 5.11A, lane 5). Interestingly, MORF4L1 knockdown does not significantly affect the expression of MRGBP (Figure 5.11A, lane 5). Similarly, MRGBP knockdown does not affect either MORF4L1, or MRFAP1, expression, as detected by immunoblot (Figure 5.11A, lane 6). Statistical analysis of the western blots indicates the statistical significance of the knockdown effects. It is noted however, that an analysis of the immunoblot signal intensity detected using the software was not performed, therefore it was not proven that all immunoblot signal intensities were detected within a linear range (Figure 5.11B).

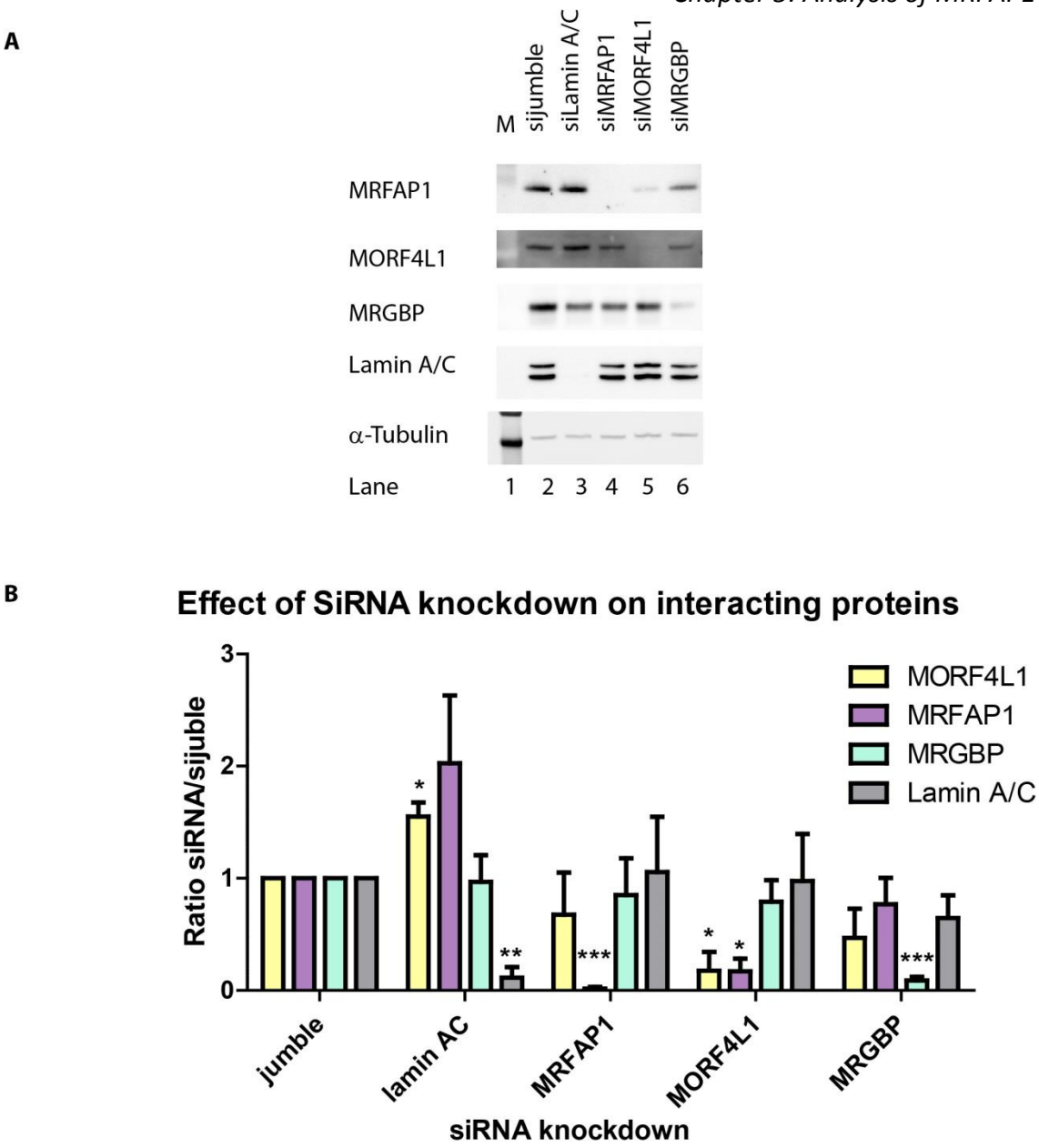


Figure 5.11. MRFAP1 expression is stabilised by MORF4L1.

A, immunoblotting of total cell lysates from U2OS cells treated with the indicated siRNA for 48 hours (n=3). 10μg of total protein loaded per lane. *B*, bar graph demonstrating the statistical analysis of the three biological replicates. The ratio of each siRNA knockdown normalised against sijumble (y-axis) is plotted for each siRNA knockdown (x-axis). Error bars indicate the standard deviation from the mean. (*=p<0.05, **=p<0.02, ***=p<0.01).

5.2.5 MRFAP1 and MORF4L1 interact as part of a small complex, distinct from the NuA4 complex

To investigate if the MRFAP1 and MORF4L1 interaction was part of known, large complexes, I interrogated the native SEC data from chapter 3, specifically looking at MRGBP as part of the NuA4 complex. This demonstrated that MRFAP1 and MRGBP coincide with the MORF4L1 elution profile in distinct and non-overlapping peaks at very different size ranges. The elution profile of MRGBP peaked in fraction 12 (Figure 5.12B) at a molecular weight of 670kDa, which correlates with previously published data on the size of the NuA4 complex [Allard et al., 1999]. In contrast, the elution profile of MRFAP1 peaked exclusively in fractions 26-28, which corresponds to a molecular weight range between 15 and 67kDa (Figure 5.12C), with no peptides detected in the high molecular weight range. These data indicate the presence of at least two distinct complexes containing MORF4L1, of very different sizes with mutually exclusive interactions with either MRGBP, or MRFAP1, respectively. To complement the mass-spectrometry-based analysis, the pairs of combined SEC fractions were immunoblotted with antibodies for MORF4L1, MRGBP and MRFAP1 (Figure 5.12D). The results from the immunoblots closely correlated with the LC-MS/MS peptide data, although MRGBP was also detected in immunoblotted fractions 12 and 13. These bands likely represented monomeric MRGBP (37kDa), as a complex containing MRGBP, MORF4L1 and MRFAP1 would be predicted to elute in a higher molecular weight range. It was noted that MRFAP1 was not detected in fractions with a molecular weight above 67kDa; therefore, it was unlikely that MRFAP1 formed part of the NuA4 complex, or indeed any other large complex contacting MORF4L1. The SEC fractionation data are thus consistent with the existence of distinct MORF4L1 complexes containing either MRGBP, or MRFAP1, as mutually exclusive interactive

partners, as judged both by MS and immunoblotting analysis to detect the elution profiles of the respective proteins.

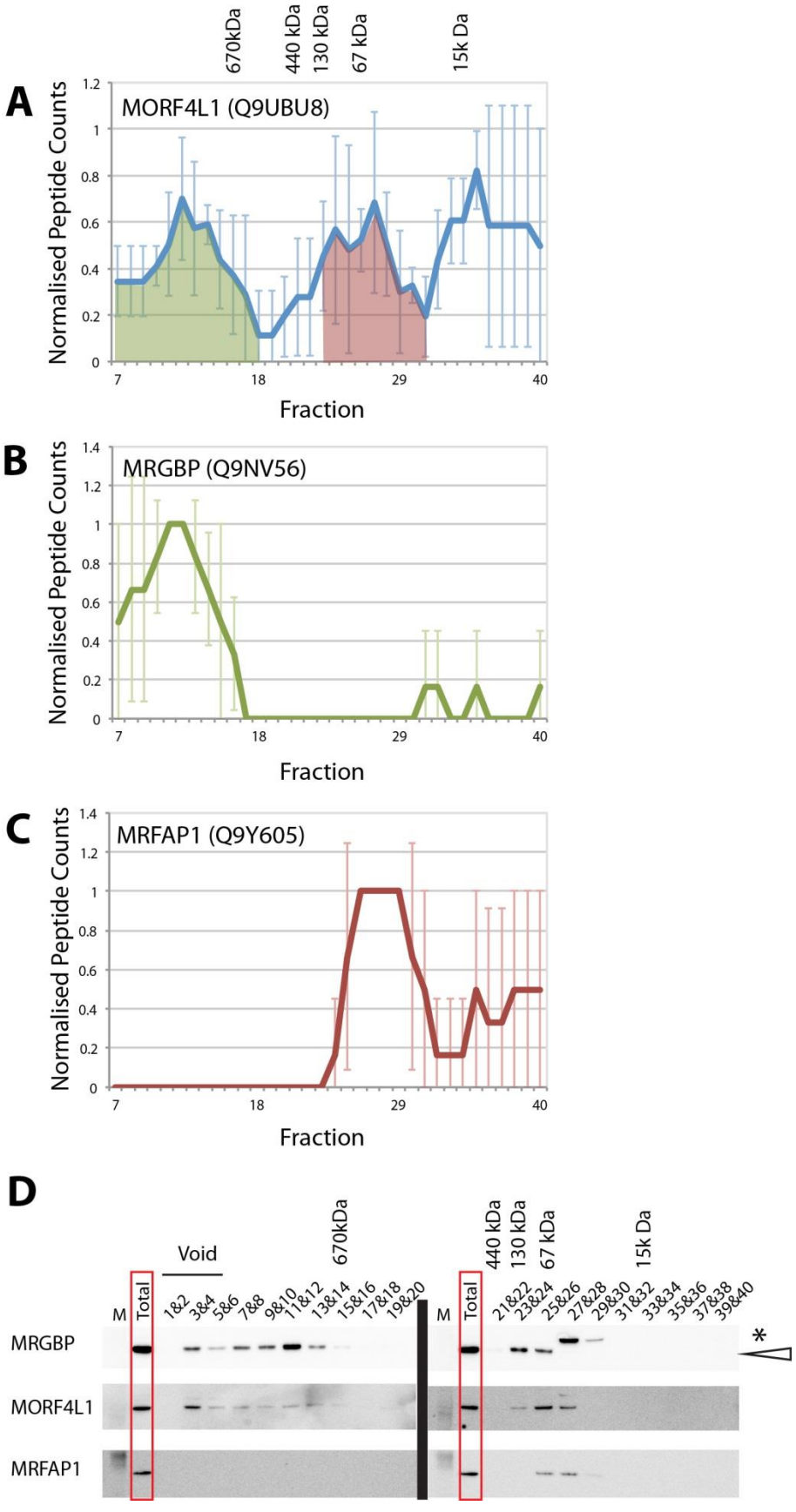


Figure 5.12. MORF4L1 interacts with MRGBP and MRFAP1 as part of two distinct complexes.

The native SEC elution profiles for MORF4L1 (*A*), MRGBP (*B*) and MRFAP1 (*C*) are shown as line graphs with normalised peptide counts (y-axis) plotted for each SEC fraction (x-axis). Peaks where MORF4L1 and MRGBP overlap are shaded in green and those where MORF4L1 and MRFAP1 overlap are in red. Error bars indicate the standard deviation from the mean. *D*, consecutive elution fractions were combined for immunoblotting. 10μg of protein was loaded in the lane labelled "total", representing the material injected onto the SEC column. A maximum of 10μg of protein was loaded in the remaining lanes. The open arrow indicates the MRGBP protein band whereas the asterisk denotes either a modified form, or, a non-specific band.

To test this further, I next performed IP experiments to examine whether MRGBP could be detected as a component that co-purifies in a pull down assay with the MRFAP1-MORF4L1 complex. To do this, the wild type, doxycycline inducible, LAPI-MRFAP1 cell line was utilised [Larance et al., 2012] to co-immunoprecipitate proteins interacting with MRFAP1. The same immunoprecipitation assay was performed on lysates from the same cells treated with MLN4924 (1μM for 16h) and MG132 (10μM for 2h). Immunoblotting of these precipitates with an antibody raised against MORF4L1 confirmed that after doxycycline induction, MORF4L1 was recovered together with GFP-MRFAP1. It also showed that this co-purification was increased when cells had been treated with either MLN4924 or MG132 (Figure 5.13). Additional MORF4L1 bands were detected by immunoblotting, at higher molecular weights than the unmodified protein. These were most marked in the lysates treated with MLN4924 and MG132 and are consistent with ubiquitinated forms of the protein. In contrast there were no specific bands detected in the EGFP-MRFAP1 immunoprecipitates that could be detected using the MRGBP antibody. These data therefore confirm the results from both the MS and protein blotting analyses of elution profiles in the SEC dataset. Collectively, the data demonstrate that MRGBP-MORF4L1 and MRFAP1-MORF4L1

complexes represent distinct functional units containing mutually exclusive interaction partners. These data also validate the use of the native SEC approach to profile co-fractionation of proteins at a system-wide level and thus to predict the existence of distinct forms of protein complexes.

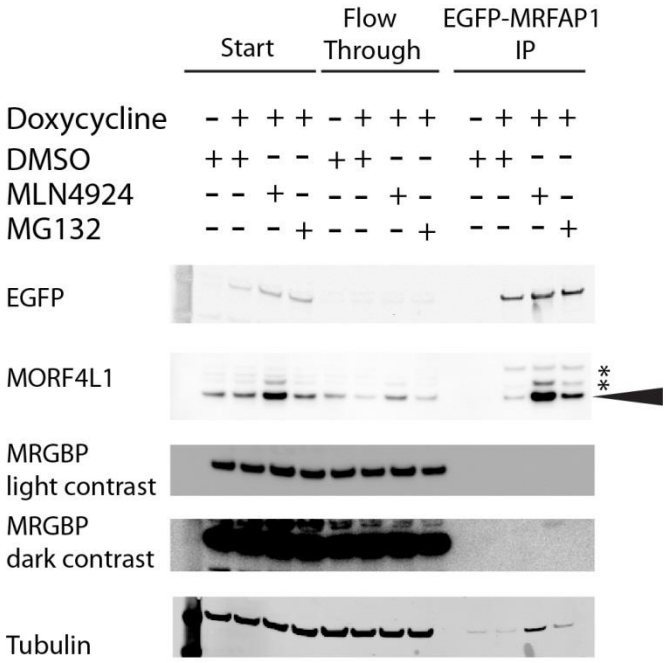


Figure 5.13. Confirmation of MORF4L1/MRGBP/MRFAP1 interactions.

Immunoblots for GFP-IP analysis of GFP-MRFAP1 cell line (n=3). Cells were induced with 10μg/ml doxycycline and treated with 1μM MLN4924 for 16 hours where indicated, 10μM MG132 where indicated, or 10μl DMSO as a control. The solid arrow indicates the unmodified MORF4L1 protein band. The asterisks indicate possible ubiquitinated forms of the MORF4L1 protein.

5.2.6 MLN4924 treatment may alter MRFAP1 interaction partners

To investigate the effect of MLN4924 treatment on protein complexes, U2OS cells were treated with 1μM MLN4924 for 20 hours prior to PBS lysis and sample preparation for SEC as previously. Half of the eluted fractions were prepared for immunoblotting and the remaining samples were digested and prepared for LC-MS/MS

analysis as previously. The lysates were immunoblotted for MRFAP1 and MORF4L1 as these proteins are known to be up-regulated in response to MLN4924 treatment. Their overall elution profiles remained unchanged in comparison to cells not subjected to the compound (Figure 5.14A). Similarly, no striking changes were seen in the immunoblots for MRGBP, tubulin and Histone H3. On reviewing the peptide elution profiles from the MS data, no significant change was seen between the untreated and MLN4924 treated cells in the MORF4L1 profile (Figure 5.14B,C). However, the MRFAP1 MLN4924 treated profile shows peptide identifications in fractions eluting earlier from the column than untreated cells, with prominent peaks in fractions 13 and 18 corresponding to complexes >670kDa and between 440 and 670kDa in size, respectively (Figure 5.14B,C). These results raise the possibility that MRFAP1 forms larger protein complexes in the presence of MLN4924.

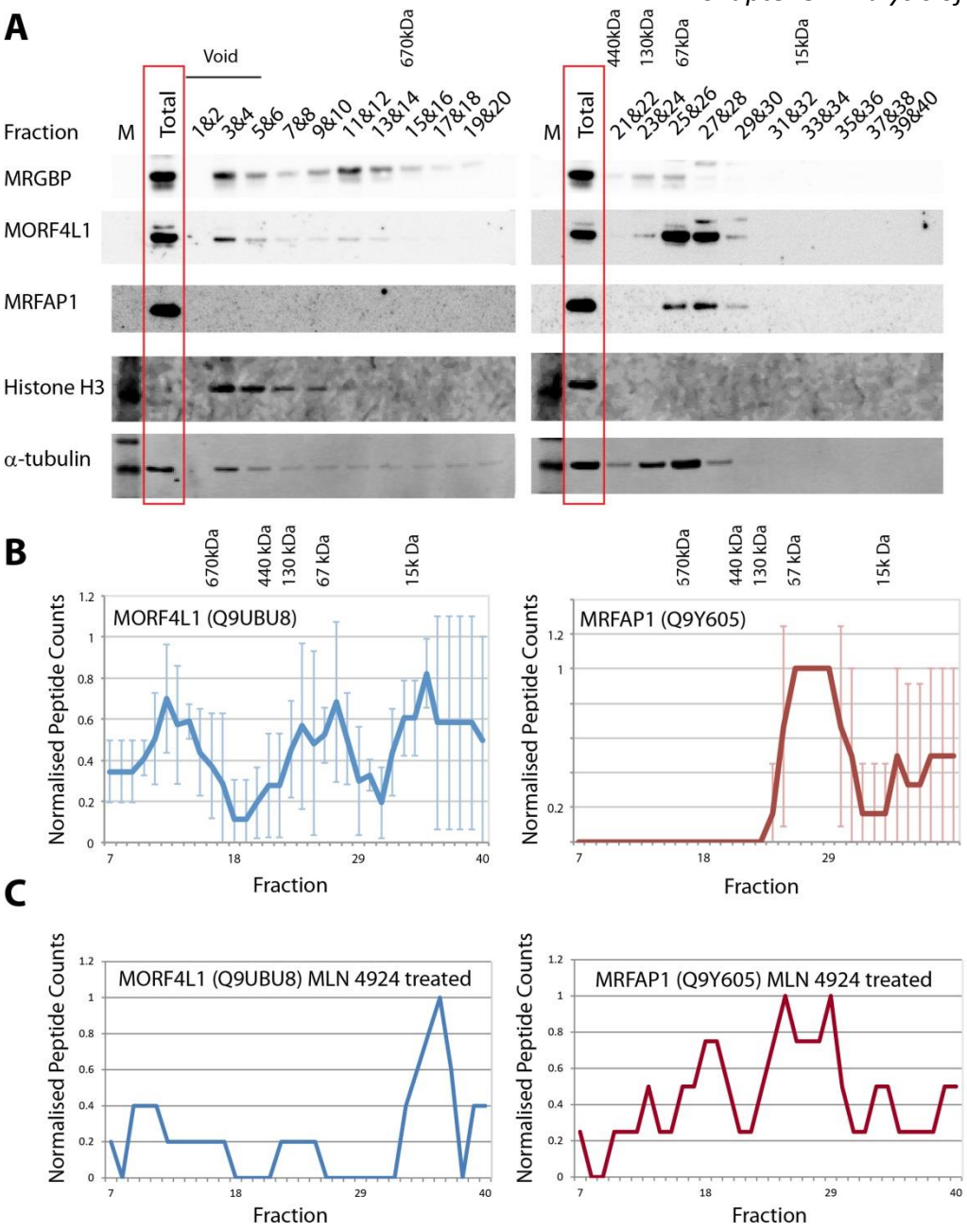


Figure 5.14. The MRFAP1 peptide elution profile changes after NEDD8 inhibition.

A, U2OS cells were treated with MLN4924 for 20 hours prior to native SEC separation. Consecutive elution fractions were combined for immunoblotting. 10μg of protein was loaded in the lane labelled "total", representing the material injected onto the SEC column. *B*, the native SEC elution profiles for MORF4L1 and MRFAP1 are shown as line graphs with normalised peptide counts (y-axis) plotted for each SEC fraction (x-axis) (n=3). Error bars indicate the standard deviation from the mean. *C*, the native SEC elution profiles for MORF4L1 and MRFAP1 after 20 hours of MLN4924 treatment are shown as line graphs with normalised peptide counts (y-axis) plotted for each SEC fraction (x-axis) (n=1).

5.2.7 MRFAP1 interacts with VprBP in the presence of MLN4924

As cullins are one of the main substrates of NEDD8, a member of this family was deemed likely to play a role in MRFAP1 regulation. CUL4B and VprBP, a known CUL4B substrate-binding protein [Le Rouzic et al., 2007; Wen et al., 2007], were identified as putative interaction partners of MRFAP1 in a previous SILAC GFP-MRFAP1 IP [Larance et al., 2012]. A commercially available antibody was used to immunoprecipitate endogenous VprBP from a total cell lysate of U2OS cells treated with MLN4924 for 18 hours. Immunoblotting the immunoprecipitate for CUL4B confirmed co-immunoprecipitation of a known VprBP interaction partner (Figure 5.15, lane 8). In addition, immunoblotting the immunoprecipitate for MRFAP1 confirmed an endogenous interaction between these two proteins as shown by the detection of a band corresponding to the MRFAP1 protein in the VprBP immunoprecipitate, with no protein detected in the rabbit IgG control immunoprecipitate (Figure 5.15, lanes 7&8).

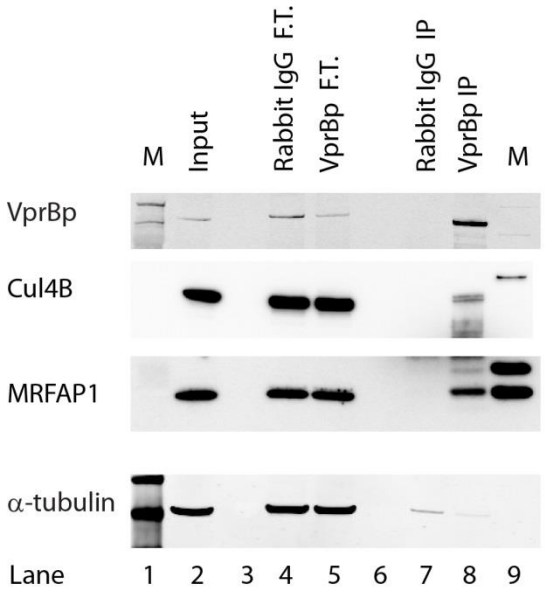


Figure 5.15. Endogenous VprBP binds MRFAP1.

3μg of VprBP antibody was used to co-immunoprecipitate VprBP interacting proteins from U2OS cells treated with 1μM MLN4924 for 16 hours. 10μg of protein was loaded per lane for the inputs and flow throughs. 10% of the immunoprecipitates were loaded in lanes 7 and 8 (n=2).

To investigate whether the additional high molecular weight peptide peaks for MRFAP1 in MLN4924 treated cells correspond to an interaction with VprBP and CUL4B, I interrogated the SEC data for these proteins in the presence of MLN4924. The elution profiles show an overlap for VprBP and MRFAP1 in fractions 12-14 but there is no clear overlap for CUL4B in these fractions (Figure 5.16). This raises the possibility that the MRFAP1 peak in this region represents the assembly of a complex containing VprBP and MRFAP1, but not CUL4B, in the setting of NEDD8 activating enzyme inhibition (Figure 5.16, grey box). However, these results would need to be investigated further through biological replicates and additional immunoprecipitation experiments to confirm a possible interaction.

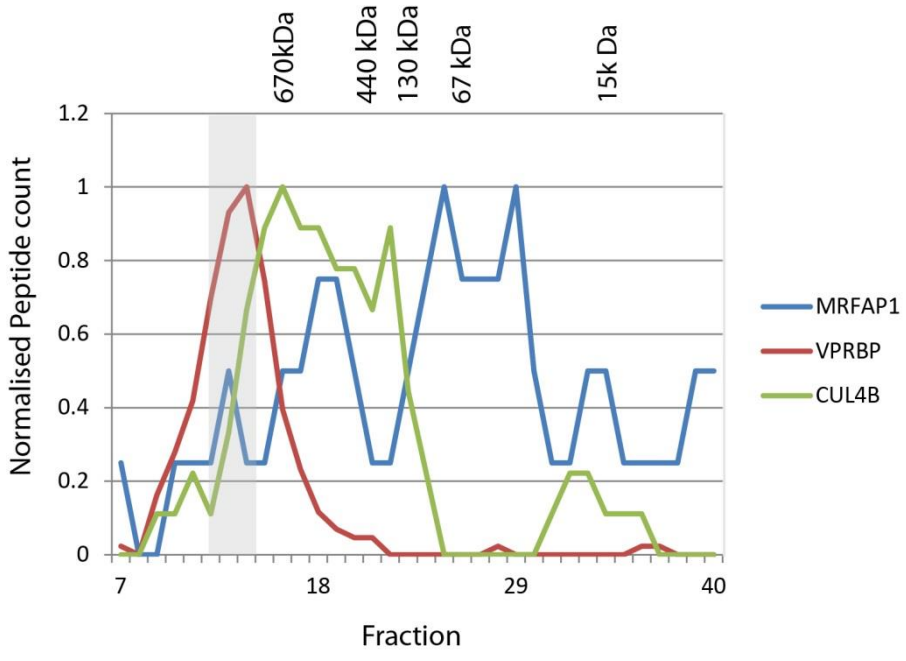


Figure 5.16. SEC elution profiles for VprBP interacting proteins in the presence of MLN4924.

Line graphs of the native SEC elution profiles for the specified proteins in the lysate of U2OS cells treated with MLN4924 for 20 hours. Normalised peptide counts (y-axis) were plotted for each fraction (x-axis). The grey box indicates the fractions in which peaks for MRFAP1 and VprBP proteins overlap.

5.2.8 MRFAP1 is expressed in spermatogonia

To investigate the possible function of the MRFAP1-MORF4L1 complex, I wanted to study the expression profile of these proteins in human tissues. The same validated antibody for the MRFAP1 protein was used by my collaborators in the Human Protein Atlas Project (www.proteinatlas.org) to screen tissue micro-arrays of both normal and malignant human tissues for the expression of the MRFAP1 protein. Analysis of these results show that the MRFAP1 protein is not widely expressed, with strongest staining identified in neuronal cells (cytoplasmic), ciliated epithelial cells of the bronchus and fallopian tubes (cytoplasmic and nuclear) and spermatogonia of the testis (nuclear) (Figure 5.17).

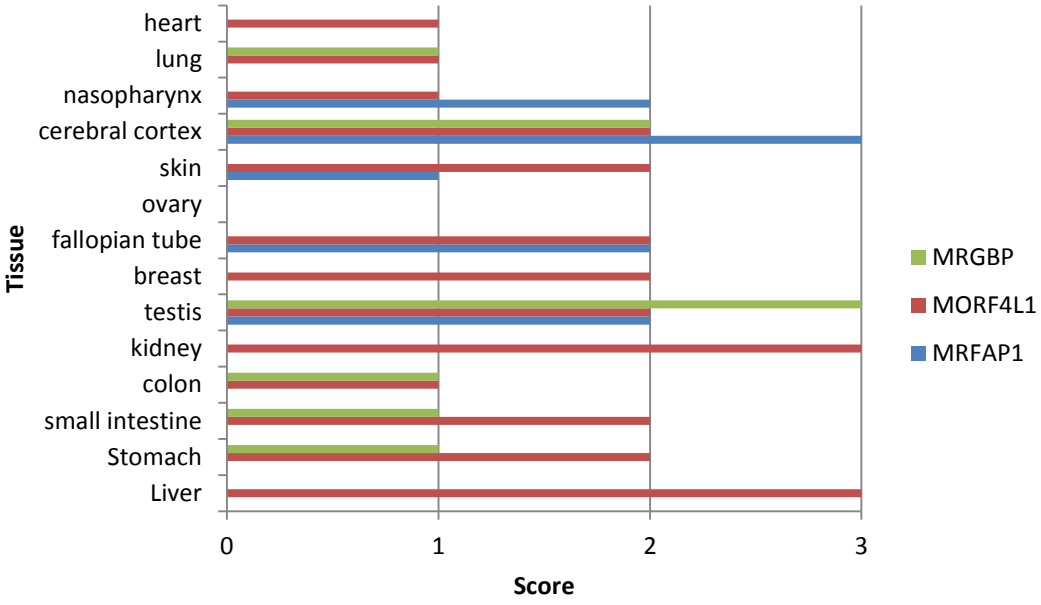


Figure 5.17. MRFAP1, MORF4L1 and MRGBP are expressed in human tissues.

Graph demonstrating the expression of MRFAP1, MORF4L1 and MRGBP in human tissues. Scoring was based on the strength of immunohistochemical staining of histological sections (0=no staining, 1=weak, 2=moderate, 3= strong staining). Data obtained from the Human Protein Atlas Project (www.proteinatlas.org) [Uhlén et al., 2005].

Interestingly, MRFAP1 was also noted to be expressed in a small number of malignancies including testicular tumours and glioblastomas. Analysis of the HPA data for MORF4L1 also demonstrated positive staining within the nuclei of spermatogenic cells of the testis, with an expression pattern that overlapped with MRFAP1 expression. MRGBP was also expressed in testis but in different cell subtypes to those expressing MRFAP1. MRGBP was expressed in occasional spermatocytes, but predominantly spermatids (Figure 5.18)

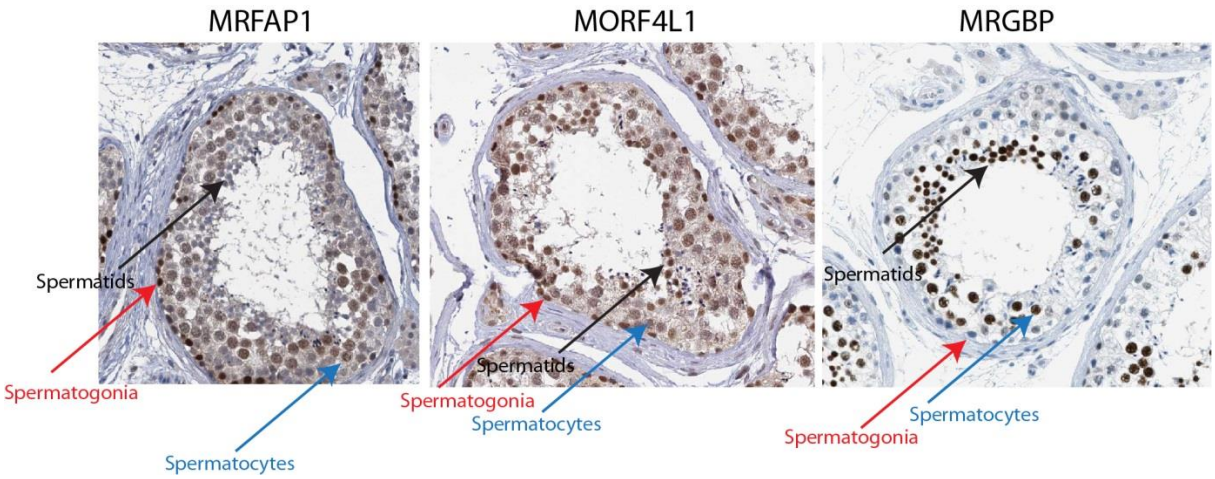


Figure 5.18. MRFAP1, MORF4L1 and MRGBP are all expressed in human testis.

Immunohistochemical staining of human testes for the stated antibodies. Spermatids (black arrows), spermatogonia (red arrows) and spermatocytes (blue arrows) are identified in each image. The brown colour indicates positive staining. Images courtesy of the Human Protein Atlas Project (www.proteinatlas.org) [Uhlén et al., 2005].

These data provide possible physiological evidence in support of the mutually exclusive MORF4L1-containing complexes, involving MRFAP1 and MRGBP, within one organ system. I therefore proceeded to use immunofluorescence microscopy to study the

expression pattern of MRFAP1, MORF4L1 and MRGBP in human testes from archived, formalin fixed, paraffin embedded resection specimens.

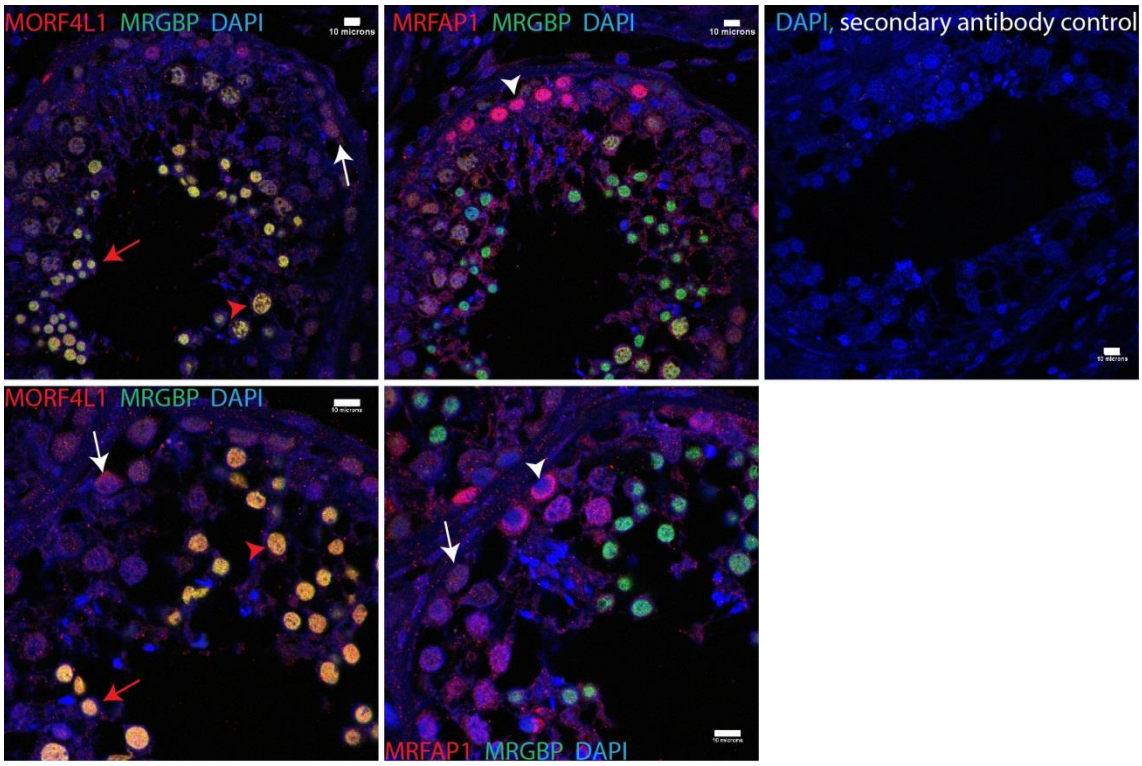


Figure 5.19. Expression of MRFAP1 in normal human testes.

Immunofluorescence microscopy of human testicular tissue. Strong nuclear staining for MRFAP1 is identified in spermatogonia (white arrowheads). MRFAP1 and MORF4L1 are both expressed in spermatogonia (white arrows). MRGBP and MORF4L1 co-localise in the nuclei of spermatocytes (red arrowheads) and spermatids (red arrows). Scale bars represent 10μm.

The sections showed that MRFAP1 expression was strongest in the nuclei of peripheral spermatogonia and spermatocytes, but showed no staining in elongated spermatids (Figure 5.19). MORF4L1 also shows positive nuclear staining in the same cell types and it is also present in spermatids. MRGBP, however, shows only minimal co-localisation with MRFAP1 in some spermatocytes, with strongest expression in spermatids, where it appears to co-localise with MORF4L1.

Analysis of human seminomas, a testicular cancer derived from malignant transformation of germ cells (spermatogonia) [Mitchell et al., 2014], shows expression of MRFAP1 in the malignant cells but minimal expression of MRGBP (Figure 5.20). These prominent staining patterns seen in the testis raise the possibility that the mutually exclusive interaction of MRFAP1 and MRGBP for binding MORF4L1 relates to specific, distinct biological roles in spermatogenesis.

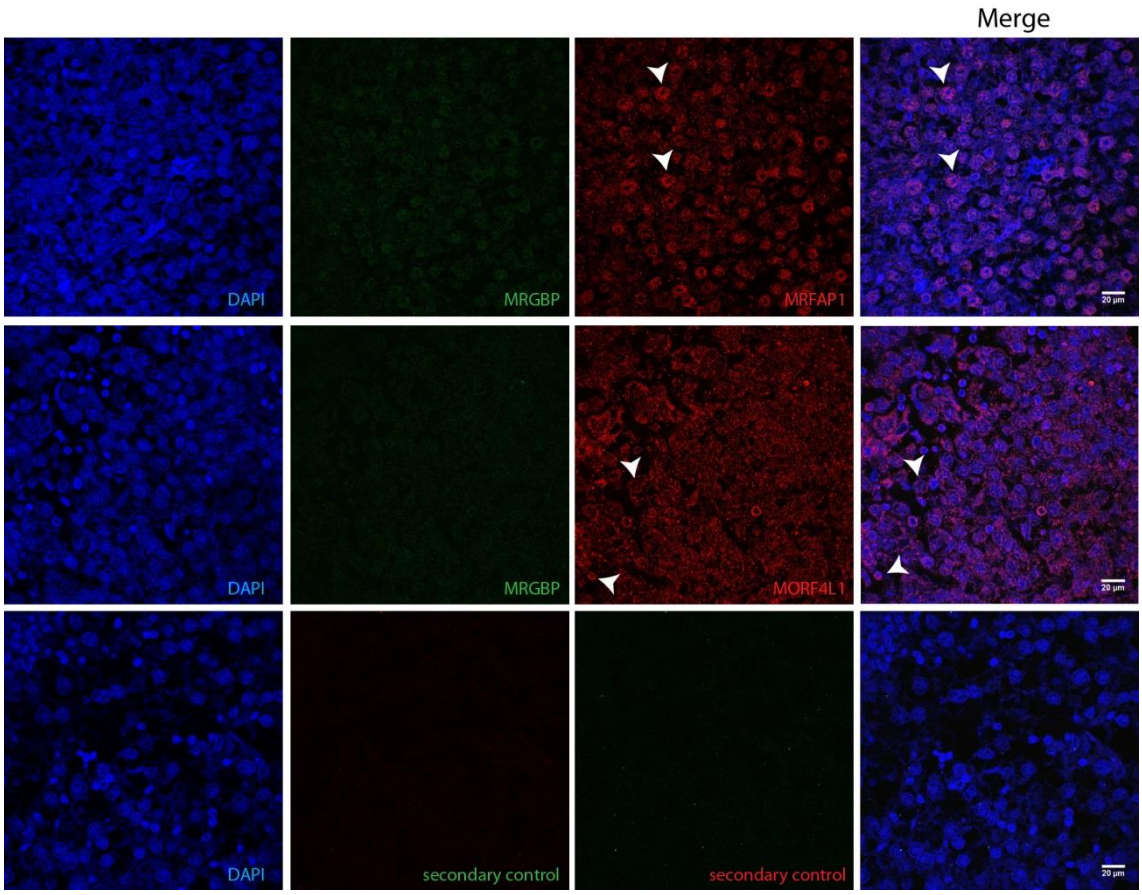


Figure 5.20. MRFAP1 is expressed in seminoma cells.

Immunofluorescence microscopy of human testicular seminoma. Nuclear positivity for MRFAP1 and MORF4L1 is identified in seminoma cells (white arrowheads). Scale bar represents 20μm.

5.2.9 Altered MRFAP1 expression may displace known interactors of the MRG domain

In view of the different expression patters of MRFAP1 and MRGBP in testis, and their interaction with MORF4L1 as part of separate complexes, I wanted to investigate whether there was competition between these two proteins for MRG domain binding. To do this, I expressed and purified the MRG domain, tagged with GST, in bacteria and coupled this to sepharose beads (Figure 5.21A,B). GST-empty vector was also expressed and bound to beads as a control. To overexpress MRFAP1 in isolation, I utilised the LAPI-7KR MRFAP1 cell line. This is a U2OS derived cell line expressing a doxycycline inducible GFP-tagged MRFAP1 mutant construct in which all seven lysine residues were converted to arginine. This rendered the MRFAP1 protein more resistant to ubiquitination. Cell lysate from un-induced LAP1-7KR cells was added to 40µl of 1mg/ml GST-MRG beads in varying concentrations to establish the saturation point for endogenous MRFAP1-MRG binding (Figure 5.21C). These results show endogenous MRFAP1 in the flow through when the pull-down was performed using 4mg total protein cell lysate, indicating that saturation of the MRG domain for MRFAP1 binding had occurred.

The beads were then used to immunoprecipitate MRG domain-binding proteins from cell lysates, both with and without MRFAP1 overexpression, to investigate whether increasing free MRFAP1 had an effect on MRGBP-MRG co-immunoprecipitation. Immunoblots demonstrate that MRFAP1 overexpression did increase the MRG-MRFAP1 interaction markedly. However there was no significant reduction in the band intensity for MRGBP in the immunoprecipitate to suggest MRFAP1 had displaced MRGBP from the MRG domain (Figure 5.22, left panel, lanes 10&11). To assess the

converse, an experiment using siRNA knockdown of the MRFAP1 protein and non-coding "jumbled" siRNA was performed in U2Os cells, and the MRG domain was again immunoprecipitated using the beads. Immunoblots show that there was no demonstrable change in MRGBP co-immunoprecipitated following MRFAP1 knockdown suggesting that, at least under these experimental conditions, there is no competitive interaction between MRFAP1 and MRGBP for the MRG binding domain of MORF4L1 (Figure 5.22, right panel, lanes 8&9).

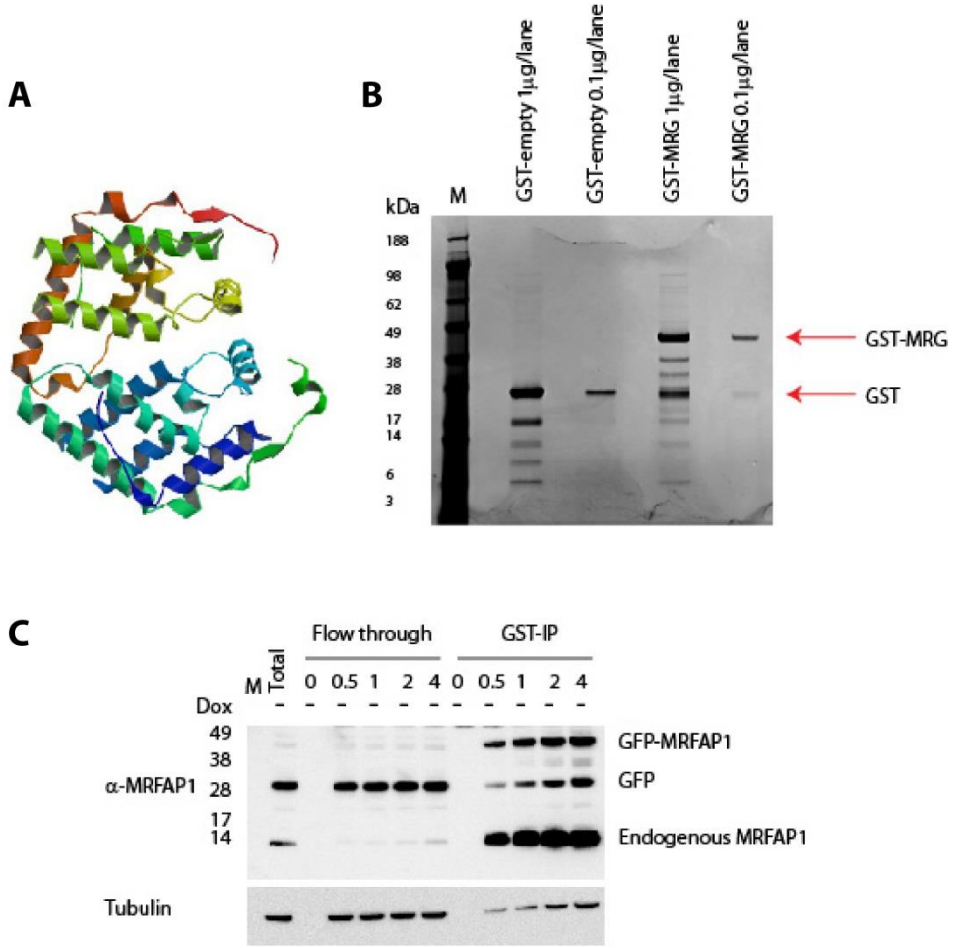


Figure 5.21. Generation of GST-MRG.

A, the MRG domain previously crystallised by Bowman *et al.* [Bowman et al., 2006] was fused to GST and expressed in *E.coli*. *B*, SYPRO ruby protein stain of the expressed GST-tagged protein and empty vector. *C*, Immunoblots of the GST-IP performed in LAP1-7KR MRFAP1 cells. 10µg of protein was loaded per lane for the flow through and 10% of the immunprecipitate was loaded in the GST-IP lanes.

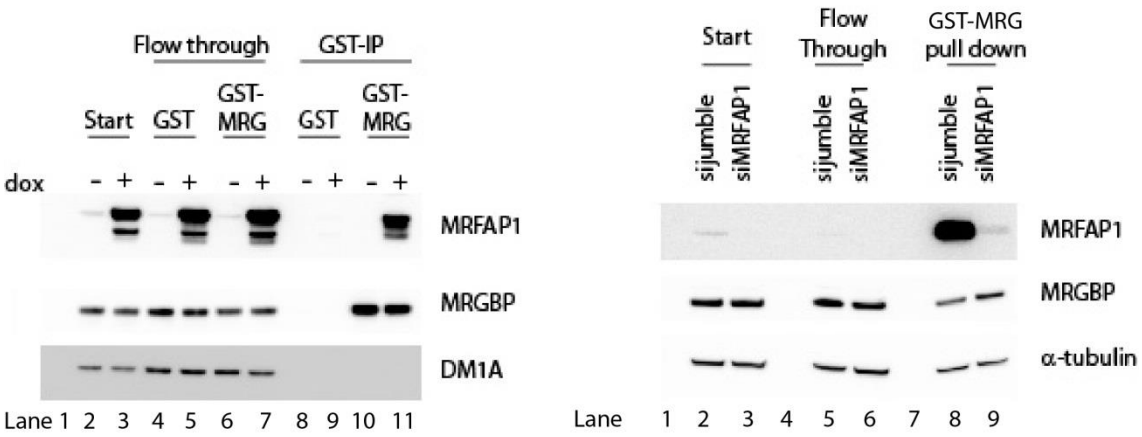


Figure 5.22. Analysis of MRFAP1 and MRGBP binding of the MRG domain.

A, immunoblots of GST-IPs performed in LAP1-7KR MRFAP1 cells either with, or without 10μg/ml doxycycline induction for 16 hours (n=2). *B*, immunoblots of GST-IPs performed in U2OS cells treated with the indicated siRNA for 48 hours. 10μg of total protein was loaded per lane for the start and flow through and 10% of the immunoprecipitate was loaded in IP lanes (n=2).

A triple SILAC pull-down experiment to assess changes in the proteins binding to the beads between uninduced 7KR-LAP1 MRFAP1 cells (light) versus doxycycline induced (heavy) cells was then performed as an unbiased approach to identify any potential competitive interaction between MRFAP1 and other MRG-domain binding proteins. The experimental design is described in Figure 5.23.

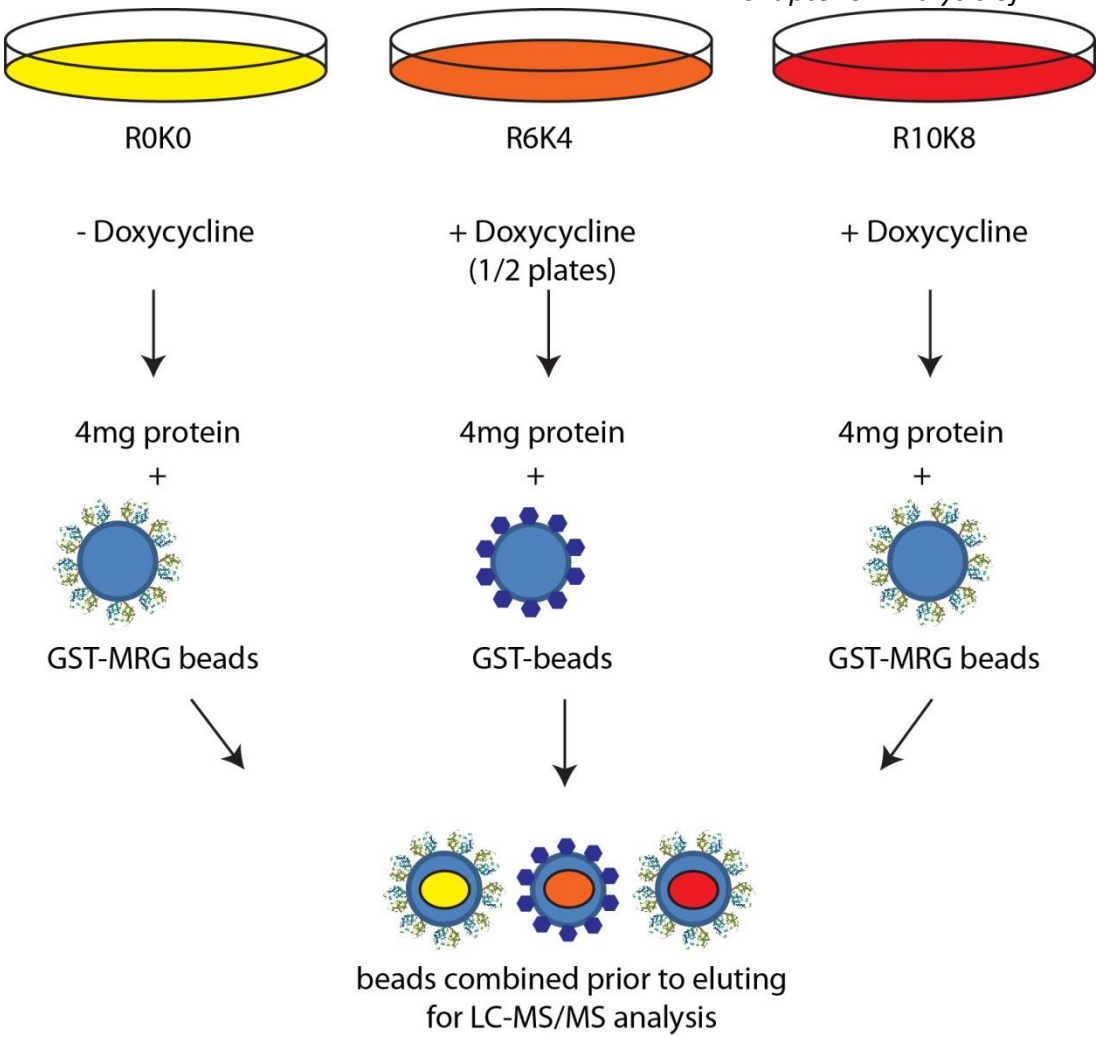
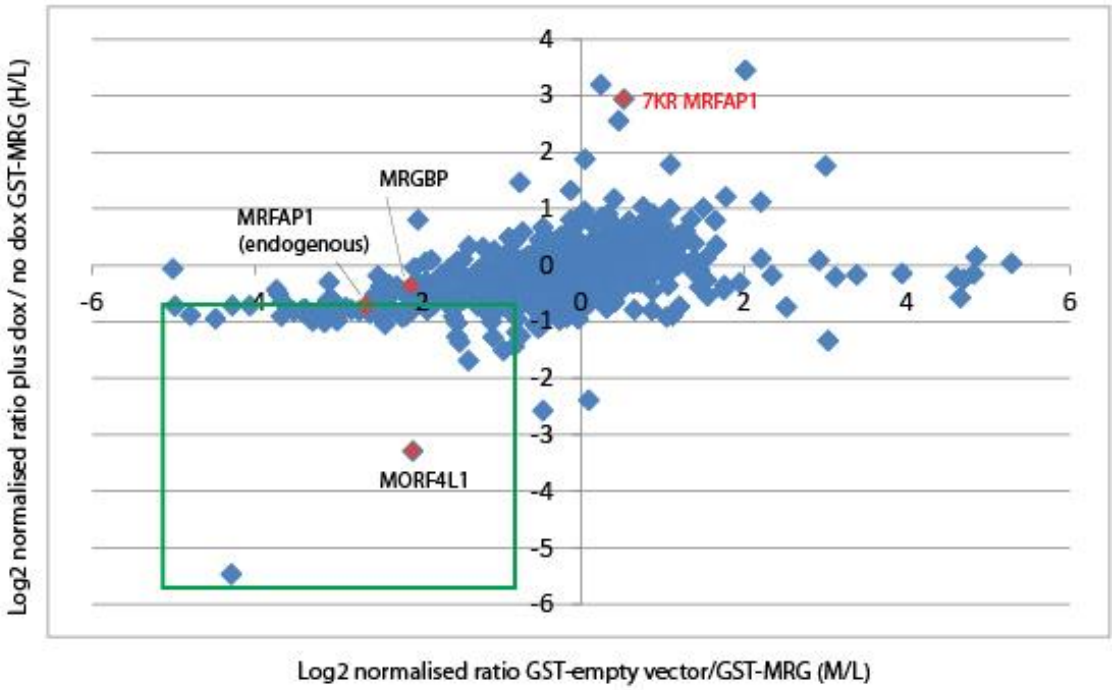


Figure 5.23. Experimental design of triple SILAC GST-MRG analysis.

The results of the SILAC analysis shows ten proteins to have low M/L ratios (Log_2 <-1, therefore specific interactors to the MRG domain) and low H/L ratios (Log_2 <-0.9, therefore a reduction in the amount binding the MRG domain in response to MRFAP1 overexpression) (Figure 5.24A,B). Of these proteins, four (Poly (rC)-binding protein 2, Retinoic acid-induced protein 3, Heat shock protein HSP90-alpha and Signal-induced proliferation-associated protein 1) have been identified in a previous SILAC study as interactors of the MORF4L1 protein [Larance et al., 2012]. Interestingly, Sin3B was shown to be a specific interactor of the MRG domain (M/L -1.0999), with a H/L Log_2

ratio of -0.909. This protein has been identified as a MRG domain interactor in a previous study [Hayakawa et al., 2007] and has recently been shown to have a role in mitigating histone acetylation and RNA polymerase II progression within transcribed loci as part of a complex containing MORF4L1 [Jelinic et al., 2011]. These results raise the possibility that MRFAP1 may competitively bind the MRG domain of MORF4L1 to regulate histone acetylation, but these results would need to be validated on obtaining a suitable antibody raised against Sin3B.

A



B

Proteins with LOG2 (M/L) normalised <-1 and LOG2 (H/L) normalised <-0.9

Protein names	LOG2 (M/L) norm	LOG2 H/L norm
Serine/threonine-protein kinase TAO2	-4.287289777	-5.466448612
Prolactin-inducible protein	-3.069391328	-4.366704579
Mortality factor 4-like protein 1	-2.067757369	-3.284616164
Shootin-1	-1.376694972	-1.689473737
Retinoic acid-induced protein 3	-1.484348095	-1.364236324
Isoform 5 of Poly(rC)-binding protein 2	-1.072149046	-1.281316167
Heat shock protein HSP 90-alpha	-1.520272757	-1.265379247
Signal-induced proliferation-associated protein 1	-2.399577991	-1.048322692
Protein enabled homolog	-1.514037414	-1.02482474
Cytospin-B	-3.141525588	-1.016803892
BTB/POZ domain-containing adapter for CUL3-mediated RhoA degradation protein 2	-1.091540605	-1.004855617
Nuclear distribution protein nudE homolog 1;Nuclear distribution protein nudE-like 1	-3.290532899	-0.992345206
Calmodulin	-2.988160903	-0.982904762
Caveolin-1	-2.189022283	-0.923632456
A-kinase anchor protein 2	-1.559791925	-0.913080393
Paired amphipathic helix protein Sin3b	-1.099896392	-0.909363329
EF-hand domain-containing protein D2	-3.672045963	-0.903738147

Figure 5.24. Analysis of competitive binding for the MRG domain.

A, a SILAC GST-MRG-IP analysis of U2OS cells stably expressing LAP1-7KR MRFAP1. The Log_2 normalised ratio of the control (GST-empty vector/GST-MRG without doxycycline induction, M/L) (x-axis) is plotted against Log_2 GST-MRG induced/not induced (H/L) (y-axis). The mutant 7KR MRFAP1 protein is indicated in red. *B*, MRG- binding proteins with a H/L Log_2 normalised ratio of <-1 and M/L Log_2 normalised ratio of <0.9.

5.3 Discussion

In this chapter, I have been able to investigate further the interaction between the MRFAP1 and MORF4L1 proteins, replicating and validating the findings of our previous large scale SILAC experiments. I have demonstrated that MRFAP1and MORF4L1 are upregulated in response to NEDDylation inhibition secondary to MLN4924 treatment, across a range of cell types derived from a range of human organs systems. I was surprised that only a small subset of proteins were identified as being up-regulated in response to MLN4924 mediated NEDD8-activating enzyme inhibition in the initial SILAC screen. Confirmation that phosphorylation of eIF2-alpha is not up-regulated in response to MLN4924 suggests that overall protein synthesis is not inhibited by this compound. This is contrary to MG132, which has been shown to inhibit protein synthesis by activation of the unfolded protein response pathway via activation of eIF2-alpha phosphorylation [Larance et al., 2013]. Another study investigating downstream targets of the NEDDylation pathway using MLN4924 in A375 melanoma cells, found up-regulation of a larger number of proteins [Liao et al., 2011]. This may reflect differences between cell lines and/or differences in MLN4924 incubation time. None the less, Liao *et al.* also demonstrated up-regulation of MRFAP1 and MOR4L1 [Liao et al., 2011], providing increased evidential support for these proteins as downstream targets of the NEDDylation pathway.

I have demonstrated that under normal conditions, MRFAP1 is rapidly degraded by the proteasome in a matter of hours and that its expression is stabilised by proteasomal inhibition. In addition, I have also demonstrated for the first time at an endogenous level, that MRFAP1 expression is stabilised by MORF4L1 and this interaction may enhance MRFAP1 stability in the context of MLN4924 treatment. Hayakawa *et al.*

[Hayakawa et al., 2007] suggest that overexpression of MRGBP may have a role in regulating either MORF4L1 synthesis, or stability, but this was not supported by my data on endogenous protein responses.

I have also demonstrated how the SEC protein complex analysis approach could be developed to study the effect of drugs on protein complexes. By comparing the elution profiles of MRFAP1 peptides in the untreated cells with the peptide elution profile for this protein in cells treated with MLN4924, clear differences are noted, with two peaks identified in the treated cells in fractions corresponding to large protein complexes, in addition to the main MORF4L1-MRFAP1 interaction peak. It is possible that the overall increase in MRFAP1 abundance caused the change in profile. However, if this were the case, it would be expected that the MORF4L1 peptide elution profile would alter too as this protein is also upregulated by MLN4924. I have demonstrated by endogenous immunoprecipitation that MRFAP1 binds to VprBP in the presence of MLN4924. The peptide elution profiles for MRFAP1 and VprBP partially overlap in MLN4924 treated cells, indicating that a complex containing these proteins may accumulate following the inhibition of the NEDD8-activating E1 enzyme. Although the elution profile for CUL4B does not closely overlap with that for VprBP, these proteins are known to interact as part of a ubiquitin E3 ring ligase complex, which regulates the replication factor Mcm10 in response to UV irradiation [Kaur et al., 2012]. From these data, I propose a model in which MRFAP1 degradation is mediated by NEDDylation via cullin 4B and its interaction with the substrate binding protein VprBP. Although the MLN4924 SEC complex analysis has only been performed as a single replicate, and additional studies would need to be performed to validate reproducibility in the results, these data serve as a proof of principle that this technique could be used to

analyse the effect of different stimuli, or experimental parameters, on protein complexes, on a proteome wide scale. The wider application of this technique to analyse changes in protein complexes in response to either different compounds, or environmental stimuli, would be an extremely interesting avenue for future research.

Size exclusion chromatography allows the separation of individual proteins that participate in multiple complexes of different sizes and/or shapes. In this chapter, I have demonstrated this separation through the analysis of the MORF4L1 adaptor protein, which is an example of a protein that is a component of several large complexes, including the NuA4 [Doyon and Cote, 2004], Sin3A [Yochum and Ayer, 2002], and BRCA1-PALB2 complexes [Hayakawa et al., 2010]. The MRFAP1 protein was known to interact with MORF4L1 [Tominaga et al., 2004], but the resulting complex had not been characterized. In my SEC dataset, the elution profiles of the MORF4L1 and MRFAP1 proteins were seen to overlap only in the small size fractions. This showed that the interaction occurred in a small complex that was separable and clearly distinct from the larger MORF4L1-containing complexes mentioned above. From the literature, it is apparent that when analysed via immunoaffinity purification, all of the MORF4L1 complexes and interaction partners co-purified together. In this case the use of SEC provided the size resolution and pre-fractionation of the separate forms of complexes necessary to help delineate differences in their protein compositions.

Contrary to my hypothesis, I was not able to demonstrate a competitive interaction between MRFAP1 and MRGBP, or Pf1 for the MRG domain of MORF4L1, by MS, in the context of stabilised-MRFAP1 overexpression in a cell line. In fact, the *in vivo* analysis suggested that the MRG domain can bind both MRFAP1 and MRGBP simultaneously.

These data are supported by mutagenesis studies, which show that although MRFAP1 and MRGBP bind the MRG domain, their specific interaction sites do not overlap [Bowman et al., 2006; Zhang et al., 2006b]. MRG binding sites for Pf1 do overlap in part with MRFAP1. However, the pull down analysis did not show a competitive interaction between this protein and MRFAP1 for the MRG domain. Sin3A binding sites for MORF4L1 may overlap with MRFAP1 [Bowman et al., 2006; Zhang et al., 2006b] but sin3A was not identified in the MRG-IP analysis. Therefore, a competitive interaction for this protein could not be assessed by MS. Interestingly, sin3B binding of the MRG domain did decrease following MRFAP1 overexpression, raising the possibility of a competitive interaction between these proteins for the MRG domain. Sin3B has previously been identified as an interactor of the MRG domain [Hayakawa et al., 2007] and interacts with MORF4L1 as part of a HDAC complex, although it is believed that the binding of Sin3B to MORF4L1 is indirect and occurs via the Pf1 – MORF4L1 interaction [Jelinic et al., 2011]. Concurrent binding of both MRFAP1 and Pf1 may alter the Pf1-Sin3B binding domain sufficiently to impair this interaction. Further validation of this competitive interaction would be required using an antibody specific for Sin3B. However, these provisional data support a role for the MRFAP1-MORF4L1 interaction in transcription regulation because impairing the ability of a HDAC complex to form as a consequence of increasing the stability of the MRFAP1-MORF4L1 interaction, will affect chromatin re-modelling, and subsequently transcription, as a result.

Another possible role for the MRFAP1-MORF4L1 interaction relates to transcription elongation. The chromodomain of MORF4L1 binds to methylated tails of histone H3, a PTM that is very important for transcription elongation. It is possible that the MRFAP1-MORF4L1 interaction is important to protect these sites, ensuring their preservation

and availability for chromatin modifying complexes. Although my data suggest that MRFAP1 and MRGBP can bind the MRG domain simultaneously on the basis of *in vitro* analysis, this is not supported by immunoprecipitation data. However, it remains possible that in the setting of the full length MORF4L1 protein, MRFAP1-MORF4L1 binding inhibits MORF4L1 binding MRGBP and/or Sin3 homologues, thereby inhibiting the chromatin binding capacity of large chromatin modifying complexes. Repeating the competitive interaction assay with full length MORF4L1-GST, rather than just the MRG domain, would address this further.

In the absence of knowledge about the absolute levels of the MRFAP1, MORF4L1 and MRGBP proteins within a cell, it is possible that the concentration of purified MRG domain used in the *in vitro* assay was too high, thus providing an alternative explanation for why no definite competitive interaction could be identified between different proteins for the MRG domain. In the future, to investigate this further, it would be beneficial to perform a more biochemical assay, namely a direct binding assay, using recombinant MRFAP1 and another recombinant MRG-binding protein. After calibrating the interaction of the individual recombinant proteins in solution with MORF4L1 immobilised on beads, the recombinant proteins (e.g. MRFAP1, MRGBP, Pf1) could be mixed in varying concentrations in solution with known quantities of MORF4L1 protein in the solid phase. Changes in the ratio of protein binding would identify any potential competitive interaction more accurately.

As MORF4L1 is a component of multiple different complexes with potentially opposing roles in transcription regulation, greater understanding of absolute protein quantitation and stoichiometric analysis of this protein and its interaction partners could potentially go some way to addressing any possible competitive interactions. The

development of protein epitope signature tags (PrESTs) [Zeiler et al., 2012], quantification concatamer (QconCAT) [Beynon et al., 2005; Pratt et al., 2006], protein standard absolute quantification (PSAQ) [Brun et al., 2007] and absolute SILAC facilitate absolute protein quantitation [Hanke et al., 2008] in line with mass spectrometry. If MORF4L1 copy numbers per cell are vastly greater than MRGBP and MRFAP1, it could explain why the distinct complexes are identified without evidence of competition between them.

Analysis of the human tissue expression pattern of MRFAP1 shows it to be expressed in only a few cell types, including spermatogonia, ciliated epithelial cells and neurones. Closer examination of the testis expression pattern of MRFAP1 and MRGBP show an anti-correlated pattern (Figure 5.19). MRFAP1 is co-expressed with MORF4L1 in spermatogonia and spermatocytes. Spermatogonia undergo mitosis to produce both primary spermatocytes and more spermatogonia, to maintain the germ cell population. Careful regulation of DNA transcription is required at this level to maintain the genetic integrity of future spermatogenesis. Therefore, it is tempting to speculate that a MORF4L1-MRFAP1 interaction at this point may have a regulatory role in germ cell differentiation. Only after subsequent differentiation into spermatocytes and spermatids did I see MRGBP expression co-localising with MORF4L1 expression, suggesting that the recruitment of NuA4 histone acetyltransferase activity to chromatin via MORF4L1 might be important for later stages of spermiogenesis. This is because only the cells in later stages of spermatogenesis co-express MRGBP and MORF4L1 but do not express MRFAP1.

Previous studies have shown in a number of animals, including rats and mice, that hyperacetylation of histone H4 is critical for spermatogenesis [Christensen and Dixon,

1982; Christensen et al., 1984; Govin et al., 2004]. These studies showed that hyperacetylation of histone H4 leads to chromatin decondensation, which is thought to allow for histone replacement by protamines and thereby facilitate the extensive chromatin compaction needed in the small sperm nucleus (Figure 5.4). I propose that the NuA4 histone acetyltransferase complex, which contains MRGBP and BRD8, plays a critical role in spermatogenesis by mediating the hyperacetylation of histone H4. One of the other MORF4L1-interacting proteins observed in the initial SILAC study was EP400, which as part of the human NuA4 complex may also play a role in the exchange of histone H2A for histone H2A.Z during spermatogenesis [Auger et al., 2008].

The testis provides a physical representation of two different MORF4L1 complexes, suggesting that complex function and temporal protein expression in cell differentiation may regulate protein-protein interactions. In support of this is the striking protein expression pattern seen in testicular seminomas. Testicular seminomas are derived from testicular germ cells and expression of MRFAP1 is still detected in these tumours, whereas MRGBP is not. This model suggests that subtle alterations in gene expression, possibly as a result of MRFAP1-MORF4L1 interactions, may facilitate malignant transformation, thereby arresting cellular differentiation and subsequent chromatin remodelling. Tominaga *et al.* report that MRFAP1 knockout mice are fertile and healthy but it would be interesting to see what the effect of an additional DNA insult has on these animals, particularly spermatogenesis [Tominaga et al., 2004]. Of course, it is also possible that mice deficient in the MRFAP1 protein compensate for loss of function through recruitment of another protein. No reports of the effects of MLN4924 on human spermatogenesis have been reported but it would be interesting to see if this compound changes the expression profiles of MRFAP1, MORF4L1 and

MRGBP. The effect of the compound on MRFAP1 in other tissues would also be of great interest, particularly those involving rapid cell division, such as the gut epithelium and skin, as this might help to shed more light on the functional role of the MRFAP1-MORF4L1 interaction.

Figure 5.25 summarises the findings of my work investigating the MRFAP1-MORF4L1 interaction, from the regulation of its degradation by NEDDylation via a pathway involving VprBP and CUL4B, to proposals for the function of the MRFAP1-MORF4L1 interaction.

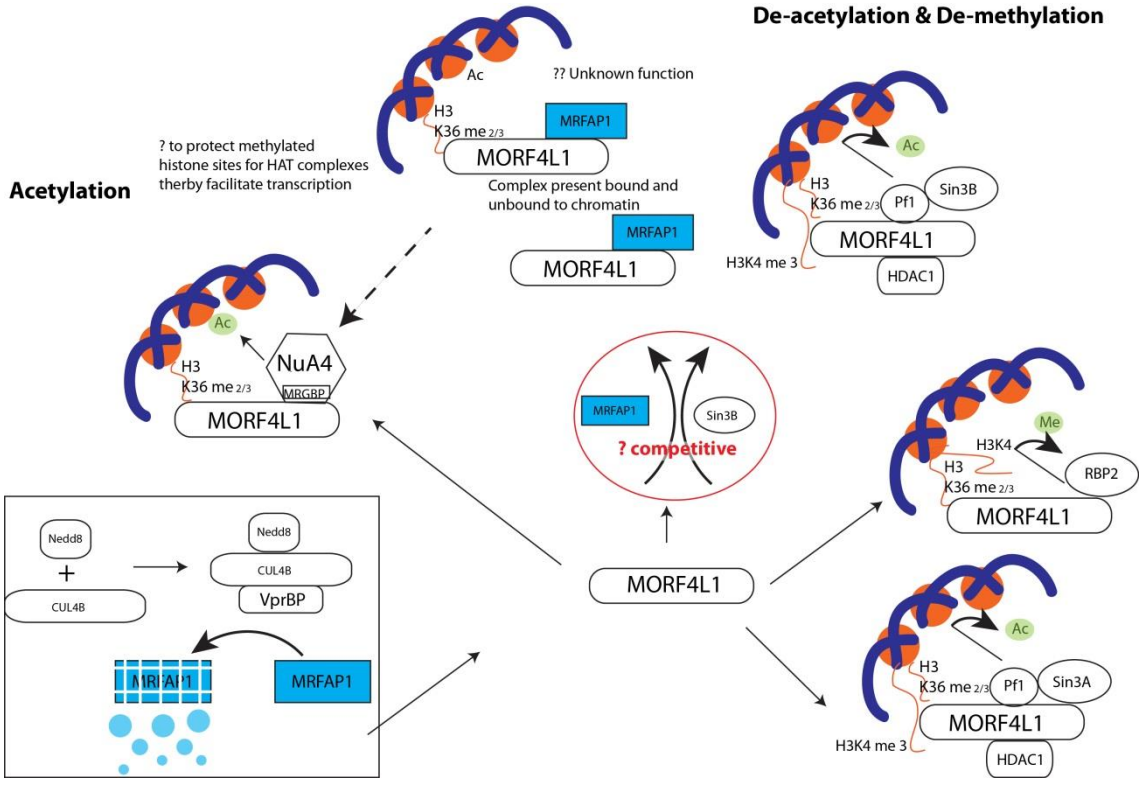


Figure 5.25. Schematic diagram of MRFAP1 and MORF4L1 interactions.

MRFAP1 is degraded by the proteome via an interaction with CUL4B and VprBP. MRFAP1 may compete with sin3B for the MRG domain of MORF4L1 and as a result, may regulate chromatin re-modelling by altering histone acetylation and deacetylation. MRFAP1 may also act to protect methylated histone sites for HAT complexes.

5.4 Distribution of effort

The mass spectrometry samples were run on the LTQ Orbitrap Velos by the "FingerPrints" proteomic facility at the College of Life Sciences, University of Dundee. Tissue sections were cut by Mr Calum Thomson of the microscopy facility, University of Dundee. Statistical analysis of the siRNA proteins in Figure 5.11 was performed in conjunction with Dr Nick Schurch of the Data Analysis Group, College of Life Sciences, University of Dundee. Data arising from the work described in this chapter contributed towards a publication [Larance et al., 2012] (Appendix II).

Chapter 6: General discussion and future directions

As I have described in the preceding chapters, proteins are the building blocks of all living things and they help to define the structure, function and regulation of cells. Many techniques have been developed to study proteins and their interactions. However, it is clear that there is a pressing need to continue to move beyond single protein identification and analysis, into the study of complex proteome dynamics. This progression is the key to a deeper understanding of cell biology.

In this thesis, I have described a strategy for the study of native protein complexes using a combination of size exclusion chromatography and tandem mass spectrometry with hierarchical clustering. I have shown using the techniques described that co-fractionating proteins can form part of known protein complexes and I have validated the results through correlation with available literature and information in public databases. Using this approach for the system-wide analysis of untagged proteins, I have also demonstrated that protein isoforms, formed either through alternative splicing, proteolytic cleavage and/or post-translational modifications, can be identified. I have provided an example for each case, demonstrating that these protein variants can be involved in different complexes from the canonical protein. These data highlight the importance of studying protein interactions at this depth and the technique described in this thesis demonstrates how this can be achieved in a reproducible, rapid and high-throughput approach.

In the analysis of the MRFAP1-MORF4L1 protein interaction, I was able to correlate immunoprecipitation data with the native SEC dataset to establish that the MORF4L1-MRFAP1 interaction formed a distinct complex to the MOR4L1-MRGBP/NuA4 complex. In addition, I have also demonstrated, through proof of principle experiments, that this approach for the study of native complexes can be used to investigate the effect of a pharmaceutical agent, such as MLN4924, on protein interactions. On the basis of these results, I propose that MRFAP1 is degraded via the proteasome through an interaction with VprBP, because in the presence of the NEDD8 activating enzyme inhibitor, higher molecular weight peaks were seen in the MRFAP1 elution profile, which corresponded to a peak in the VprBP peptide elution profile. This proposal has been supported by immunoprecipitation data which confirm an endogenous interaction between MRFAP1 and VprBP in the presence of MLN4924. Provisional results described in Appendix I also demonstrate that this technique could be applied to the analysis of protein complexes in whole tissues, which is a very exciting prospect because large scale, high-throughput comparative analyses of complexes under different conditions, e.g. starvation or drug treatment, would not be easily achievable using affinity purification methods. In addition, this type of analysis would produce more clinically relevant results as many cancer cell lines no longer accurately reflect their disease of origin.

I proceeded to develop the native protein complex analysis described in chapter three, by the addition of *in vivo* protein crosslinking, denaturing cell lysis and denaturing SEC, prior to MS/MS analysis. This development led to the more efficient recovery of proteins underrepresented in the native SEC dataset and as a result, improved my ability to resolve membrane complexes, such as integrins, many of which were not identified by the native SEC approach. It would be of great interest in the future to

perform a comparative analysis between the crosslinked dataset and the PBS dataset from chapter three, to confirm that protein crosslinking has improved the overall resolution and increased the identification of protein complexes, over and above the native approach. In addition, by analysing the protein complexes identified in lysates obtained from cells crosslinked by varying concentrations of formaldehyde, e.g. 1% and 4%, it would be useful to use these results, in combination with the native analysis, to score and validate protein interactions, and to facilitate the identification of potential novel protein interactions and protein complexes.

In the study of endogenous protein complexes, the use of a crosslinker can increase the efficiency of protein recovery. Formaldehyde has proved to be a very useful reagent for analysis of endogenous protein complexes. However, it remains rather a "blunt tool", as there is very little control over the position of inter-protein crosslinks formed. In addition, truly interacting proteins may not have been crosslinked by formaldehyde purely as a result of the linker arm being too short to bridge the reactive residues. In the future, performing the crosslinking with a reagent with a longer linker arm, such as DSP (12Å), may enable more protein complex subunits to be crosslinked, such as the α and β subunits of the proteasome.

Very limited crosslinking studies analysing protein interaction sites have been performed in whole cell lysates [Rinner et al., 2008; Xu et al., 2010; Yang et al., 2012], partly due to the complexity of the reference library of potential crosslinked peptides. As I was not studying the specific sites of protein crosslinking for structural analysis of the complexes and as I reversed crosslinks prior to MS analysis, some of the limitations of crosslinking-MS/MS approaches were not as problematic in my study. Of course, should the SEC crosslinking technique be extended to include this information in the

future, a different crosslinker would need to be selected to facilitate the enrichment of crosslinked peptides, as without enrichment, crosslinked peptides can be difficult to detect in protein digests due to low ion signals [Petrotchenko and Borchers, 2010]. For example, by utilising a crosslinker with an affinity tag within the spacer arm it is possible to enrich for crosslinked peptides by affinity purification [Hurst et al., 2004; Petrotchenko et al., 2011]. As an additional feature, isotopically labelling the crosslinker provides doublet peaks for crosslinked peptides, which can then be detected in complex mass spectra [Müller et al., 2001]. Selection of a crosslinker, which could be reliably cleaved within the spacer arm, directly in the mass spectrometer, would enable specific crosslinked peptides to be identified without the need for a crosslinked peptide reference library. Combining the affinity purification and controlled cleavage approaches would increase the detection of specific protein-protein interactions and identify their crosslinking sites. This combined approach was used by Petrotchenko *et al.* to analyse the structure of the protein heterodimer of HIV reverse transcriptase using the crosslinker, cyanurbiotindipropionylsuccinimide (CBDPS) [Petrotchenko et al., 2011]. However, the use of combined isotopic labelling, affinity tagging and directed linker cleavage in mass spectrometry has not yet been applied to analyse complex structures at a proteome wide level and this would be a very exciting avenue for future study.

In addition to the development of new crosslinking agents to improve this methodology even further, ongoing development of SEC column technology, such as modifications to column design and the development of new silica matrices, will also improve our ability to resolve protein complexes even further. A significant advantage of crosslinking protein complexes is that the interactions are not perturbed by

different buffer conditions. Havugimana et.al. developed an approach for the study of soluble protein complexes, which involved a combination of ion exchange, isoelectric focusing and sucrose gradient separation [Havugimana et al., 2012]. The multi-pronged approach is a significant strength of this analysis. However, the different buffer conditions across the techniques may have resulted in some protein interactions not being replicated across the techniques. It would be an exciting prospect to analyse crosslinked proteins in a 2 dimensional approach, first separating complexes by ion exchange chromatography and then separating each fraction again by size using SEC. Buffer conditions would not perturb the crosslinked proteins and this methodology would permit even further resolution of complexes. In particular, complexes which co-elute by SEC on the basis of size, without interacting, would be unlikely to have the same charge and therefore would be likely to be resolved in different fractions at the end of the 2D analysis.

For the crosslinking analysis, the method used for hierarchical clustering was modified to cluster the proteins across the 144 fractions, rather than clustering on the mean of three biological replicates, as was done for the native dataset in chapter three. As a result, more complexes could be resolved within separate clusters without breaking down known interacting proteins. In future, statistical analysis by means of ROC curve calculation, which would utilise a false positive and false negative calculation to select the optimum cluster number, will improve the value of the analysis even further. Another area, which is currently undergoing development in the Lamond laboratory, is the development of software which will help to identify more interacting proteins. This software is designed to identify peaks in an elution profile and would match proteins peaking within the same elution fraction. This development will particularly help to

identify proteins that are involved in more than one complex. Other research groups have utilised publicly available datasets to refine and validate protein interaction studies obtained from mass spectrometry experiments [Byron et al., 2012]. Applying a similar comparative analysis to identify co-fractionating proteins between the two datasets contained in this thesis, and/or comparing the results to protein-protein interactions identified in other large scale protein complex analyses [Havugimana et al., 2012], would provide a wealth of knowledge about previously unidentified protein-protein interactions.

This thesis has mainly focussed on method development. However, I have also demonstrated how the SEC/MS approach for protein complex analysis could be applied to the study of the effect of drugs on complexes and for the study of protein complexes within tissues. There are many more applications that would be ideally suited for this approach. For example, Larance et.al. used subcellular fractionation to compare protein degradation rates between subcellular compartments [Larance et al., 2013]. The crosslinking SEC method described in this thesis could be combined with subcellular fractionation to identify the localisation of protein complexes throughout the cell. Similarly, methods described by Ly et.al., in the quantitative analysis of protein variation across the cell cycle, could also be combined with the crosslinking technique [Ly et al., 2014]. This type of analysis could provide additional biologically relevant data that would not be highlighted from single protein identifications alone. For example, the total abundance of a protein in a cell may not change across the cell cycle. However, it is entirely feasible that the interaction partners, and therefore the protein complexes, that a given protein interacts with do change, and, as a result, this change will affect cellular functions. Also, the ability to analyse modified proteins and their

involvement in protein complexes will provide an additional depth to the analysis. By incorporating an additional parameter, such as a quantitative labelling technique (e.g. tandem mass tags), either in the cell cycle analysis, or across subcellular compartments, it would be possible to study protein complex assembly over different time points. My methods for the analysis of protein complexes could also extend beyond tissue culture and into the clinical setting. Individual proteins may not change in overall abundance in response to either environmental stimuli, or disease states, but protein interactions can vary, and, as a result, studying the proteome at the level of protein complexes may give a clearer picture of the intracellular landscape.

These are just a small number of examples that demonstrate just how powerful the endogenous protein complex, SEC/MS approach could be. This approach permits the analysis of an additional dimension in proteomics: the study of protein abundance, post-translational modifications and now protein interactions are all possible.

References

Adams, J. (2003) The proteasome: structure, function, and role in the cell. Cancer Treatment Reviews *29, Supplement 1(0)*: 3-9.

Aebersold, R., M. Mann (2003) Mass spectrometry-based proteomics. Nature *422(6928)*: 198-207.

Ahmad, Y., F.-M. Boisvert, E. Lundberg, M. Uhlen, A.I. Lamond (2012) Systematic Analysis of Protein Pools, Isoforms, and Modifications Affecting Turnover and Subcellular Localization. Molecular & Cellular Proteomics *11(3)*.

Allard, S., R.T. Utley, J. Savard, A. Clarke, P. Grant, C.J. Brandl, L. Pillus, J.L. Workman, J. Cote (1999) NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J *18(18)*: 5108-5119.

Aloy, P., R.B. Russell (2002) Interrogating protein interaction networks through structural biology. Proceedings of the National Academy of Sciences *99(9)*: 5896-5901.

Aloy, P., R.B. Russell (2003) InterPreTS: protein Interaction Prediction through Tertiary Structure. Bioinformatics *19(1)*: 161-162.

Alpert, A.J. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr *499*: 177-196.

Auger, A., L. Galarneau, M. Altaf, A. Nourani, Y. Doyon, R.T. Utley, D. Cronier, S. Allard, J. Côté (2008) Eaf1 Is the Platform for NuA4 Molecular Assembly That Evolutionarily Links Chromatin Acetylation to ATP-Dependent Exchange of Histone H2A Variants. Mol Cell Biol *28(7)*: 2257-2270.

Bai, Y., K. Markham, F. Chen, R. Weerasekera, J. Watts, P. Horne, Y. Wakutani, R. Bagshaw, P.M. Mathews, P.E. Fraser, D. Westaway, P. St. George-Hyslop, G. Schmitt-Ulms (2008) The in Vivo Brain Interactome of the Amyloid Precursor Protein. Molecular & Cellular Proteomics *7(1)*: 15-34.

Beck, M., A. Schmidt, J. Malmstroem, M. Claassen, A. Ori, A. Szymborska, F. Herzog, O. Rinner, J. Ellenberg, R. Aebersold (2011) The quantitative proteome of a human cell line.

Belgareh, N., G. Rabut, S.W. Baï, M. van Overbeek, J. Beaudouin, N. Daigle, O.V. Zatsepina, F. Pasteau, V. Labas, M. Fromont-Racine, J. Ellenberg, V. Doye (2001) An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells. The Journal of Cell Biology *154(6)*: 1147-1160.

Benjamini, Y., Y. Hochberg (1995) Controlling the false discovery rate: a practical and powerful approach to mulitple testing. Journal of the Royal Statistical Society Series B *57*: 289-300.

Beynon, R.J., M.K. Doherty, J.M. Pratt, S.J. Gaskell (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat Meth *2(8)*: 587-589.

Blagoev, B., S.-E. Ong, I. Kratchmarova, M. Mann (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotech *22(9)*: 1139-1145.

Bochman, M.L., A. Schwacha (2009) The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. Microbiology and Molecular Biology Reviews *73(4)*: 652-683.

Boisvert, F.-M., Y. Ahmad, M. Gierliński, F. Charrière, D. Lamont, M. Scott, G. Barton, A.I. Lamond (2012) A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. Molecular & Cellular Proteomics *11(3)*.

Bonizzi, G., M. Karin (2004) The two NF-κB activation pathways and their role in innate and adaptive immunity. Trends in Immunology *25(6)*: 280-288.

Boulon, S., Y. Ahmad, L. Trinkle-Mulcahy, C. Verheggen, A. Cobley, P. Gregor, E. Bertrand, M. Whitehorn, A.I. Lamond (2010) Establishment of a Protein Frequency Library and Its Application in the Reliable Identification of Specific Protein Interaction Partners. Molecular & Cellular Proteomics *9(5)*: 861-879.

Bowman, B.R., C.M. Moure, B.M. Kirtane, R.L. Welschhans, K. Tominaga, O.M. Pereira-Smith, F.A. Quiocho (2006) Multipurpose MRG Domain Involved in Cell Senescence and Proliferation Exhibits Structural Homology to a DNA-Interacting Domain. Structure *14(1)*: 151-158.

Bruckner, A., C. Polge, N. Lentze, D. Auerbach, U. Schlattner (2009) Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci *10(6)*: 2763-2788.

Brun, V., A. Dupuis, A. Adrait, M. Marcellin, D. Thomas, M. Court, F. Vandenesch, J. Garin (2007) Isotope-labeled Protein Standards: Toward Absolute Quantitative Proteomics. Molecular & Cellular Proteomics *6(12)*: 2139-2149.

Byron, A., J.D. Humphries, S.E. Craig, D. Knight, M.J. Humphries (2012) Proteomic analysis of α4β1 integrin adhesion complexes reveals α-subunit-dependent protein recruitment. PROTEOMICS *12(13)*: 2107-2114.

Cai, Y., J. Jin, C. Tomomori-Sato, S. Sato, I. Sorokina, T.J. Parmely, R.C. Conaway, J.W. Conaway (2003) Identification of New Subunits of the Multiprotein Mammalian TRRAP/TIP60-containing Histone Acetyltransferase Complex. Journal of Biological Chemistry *278(44)*: 42733-42736.

Camacho-Carvajal, M.M., B. Wollscheid, R. Aebersold, V. Steimle, W.W.A. Schamel (2004) Two-dimensional Blue Native/SDS Gel Electrophoresis of Multi-Protein Complexes from Whole Cellular Lysates: A PROTEOMICS APPROACH. Molecular & Cellular Proteomics *3(2)*: 176-182.

Campbell, I.D., M.J. Humphries (2011) Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol *3(3)*.

Cargile, B.J., D.L. Talley, J.L. Stephenson (2004) Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. ELECTROPHORESIS *25(6)*: 936-945.

Carter, S., O. Bischof, A. Dejean, K.H. Vousden (2007) C-terminal modifications regulate MDM2 dissociation and nuclear export of p53. Nat Cell Biol *9(4)*: 428-435.

Catherman, A.D., O.S. Skinner, N.L. Kelleher (2014) Top Down proteomics: Facts and perspectives. Biochemical and Biophysical Research Communications *445(4)*: 683-693.

Chang, I.-F. (2006) Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes. PROTEOMICS *6(23)*: 6158-6166.

Chang, I.F., K. Szick-Miranda, S. Pan, J. Bailey-Serres (2005) Proteomic characterization of evolutionarily conserved and variable proteins of Arabidopsis cytosolic ribosomes. Plant Physiol *137(3)*: 848-862.

Chen, M., M. Takano-Maruyama, O.M. Pereira-Smith, G.O. Gaufo, K. Tominaga (2009) MRG15, a component of HAT and HDAC complexes, is essential for

proliferation and differentiation of neural precursor cells. Journal of Neuroscience Research *87(7)*: 1522-1531.

Chen, Z.A., A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.C. Bukowski-Wills, M. Nilges, P. Cramer, J. Rappsilber (2010) Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. The EMBO Journal *29(4)*: 717-726.

Choi, A.M.K., S.W. Ryter, B. Levine (2013) Autophagy in Human Health and Disease. New England Journal of Medicine *368(7)*: 651-662.

Choudhary, C., C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, J.V. Olsen, M. Mann (2009) Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science *325(5942)*: 834-840.

Christensen, M.E., G.H. Dixon (1982) Hyperacetylation of histone H4 correlates with the terminal, transcriptionally inactive stages of spermatogenesis in rainbow trout. Dev Biol *93(2)*: 404-415.

Christensen, M.E., J.B. Rattner, G.H. Dixon (1984) Hyperacetylation of histone H4 promotes chromatin decondensation prior to histone replacement by protamines during spermatogenesis in rainbow trout. Nucleic Acids Research *12(11)*: 4575-4592.

Chu, F., J.C. Maynard, G. Chiosis, C.V. Nicchitta, A.L. Burlingame (2006) Identification of novel quaternary domain interactions in the Hsp90 chaperone, GRP94. Protein Science *15(6)*: 1260-1269.

Clague, M.J., S. Urbé (2010) Ubiquitin: Same Molecule, Different Degradation Pathways. Cell *143(5)*: 682-685.

Clamp, M., B. Fry, M. Kamal, X. Xie, J. Cuff, M.F. Lin, M. Kellis, K. Lindblad-Toh, E.S. Lander (2007) Distinguishing protein-coding and noncoding genes in the human genome. Proceedings of the National Academy of Sciences *104(49)*: 19428-19433.

Cohen, P. (2009) Targeting protein kinases for the development of anti-inflammatory drugs. Current Opinion in Cell Biology *21(2)*: 317-324.

Cohen, P., M. Tcherpakov (2010) Will the ubiquitin system furnish as many drug targets as protein kinases? Cell *143(5)*: 686-693.

Coll, J.M., J.W. Sekowski, R.J. Hickey, L. Schnaper, W. Yue, A.M. Brodie, L. Uitto, J.E. Syvaoja, L.H. Malkas (1996) The human breast cell DNA synthesome: its purification from tumor tissue and cell culture. Oncol Res *8(10-11)*: 435-447.

Collins, M.O., J.S. Choudhary (2008) Mapping multiprotein complexes by affinity purification and mass spectrometry. Current Opinion in Biotechnology *19(4)*: 324-330.

Consortium, T.U. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Research *41(D1)*: D43-D47.

Cowan, K.J., K.B. Storey (2003) Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress. Journal of Experimental Biology *206(7)*: 1107-1115.

Cox, J., M. Mann (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech *26(12)*: 1367-1372.

Cox, J.r., N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann (2011) Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. Journal of Proteome Research *10(4)*: 1794-1805.

Cronshaw, J.M., A.N. Krutchinsky, W. Zhang, B.T. Chait, M.J. Matunis (2002) Proteomic analysis of the mammalian nuclear pore complex. The Journal of Cell Biology *158(5)*: 915-927.

Deng, D., C. Xu, P. Sun, J. Wu, C. Yan, M. Hu, N. Yan (2014) Crystal structure of the human glucose transporter GLUT1. Nature *510(7503)*: 121-125.

Dos, D.S., S.M. Ali, D.-H. Kim, D.A. Guertin, R.R. Latek, H. Erdjument-Bromage, P. Tempst, D.M. Sabatini (2004) Rictor, a Novel Binding Partner of mTOR, Defines a Rapamycin-Insensitive and Raptor-Independent Pathway that Regulates the Cytoskeleton. Current Biology *14(14)*: 1296-1302.

Doyon, Y., J. Cote (2004) The highly conserved and multifunctional NuA4 HAT complex. Curr Opin Genet Dev *14(2)*: 147-154.

Driscoll, J.J., R.D. Chowdhury (2012) Molecular crosstalk between the proteasome, aggresomes and autophagy: Translational potential and clinical implications. Cancer letters *325(2)*: 147-154.

Droit, A., G.G. Poirier, J.M. Hunter (2005) Experimental and bioinformatic approaches for interrogating protein–protein interactions to determine protein function. Journal of Molecular Endocrinology *34(2)*: 263-280.

Ethier, M., J.-P. Lambert, J. Vasilescu, D. Figeys (2006) Analysis of protein interaction networks using mass spectrometry compatible techniques. Analytica Chimica Acta *564(1)*: 10-18.

Eubel, H., H.-P. Braun, A.H. Millar (2005) Blue-native PAGE in plants: a tool in analysis of protein-protein interactions. Plant Methods *1(1)*: 11.

Ewing, R.M., P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M.D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y.V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J.-P. Lambert, H.S. Duewel, I.I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.-L. Adams, M.F. Moran, G.B. Morin, T. Topaloglou, D. Figeys (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol *3*.

Fasshauer, D. (2003) Structural insights into the SNARE mechanism. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1641(2–3)*: 87-97.

Fenczik, C.A., R. Zent, M. Dellos, D.A. Calderwood, J. Satriano, C. Kelly, M.H. Ginsberg (2001) Distinct Domains of CD98hc Regulate Integrins and Amino Acid Transport. Journal of Biological Chemistry *276(12)*: 8746-8752.

Feng, Y., D. He, Z. Yao, D.J. Klionsky (2014) The machinery of macroautophagy. Cell Res *24(1)*: 24-41.

Fontoura, B.M.A., G. Blobel, M.J. Matunis (1999) A Conserved Biogenesis Pathway for Nucleoporins: Proteolytic Processing of a 186-Kilodalton Precursor Generates Nup98 and the Novel Nucleoporin, Nup96. The Journal of Cell Biology *144(6)*: 1097-1112.

Gambus, A., G.A. Khoudoli, R.C. Jones, J.J. Blow (2011) MCM2-7 Form Double Hexamers at Licensed Origins in Xenopus Egg Extract. Journal of Biological Chemistry *286(13)*: 11855-11864.

Garman, E.F. (2014) Developments in x-ray crystallographic structure determination of biological macromolecules. Science *343(6175)*: 1102-1108.

Gaucher, J., N. Reynoird, E. Montellier, F. Boussouar, S. Rousseaux, S. Khochbin (2010) From meiosis to postmeiotic events: the secrets of histone disappearance. Febs J *277(3)*: 599-604.

Gavin, A.-C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L.J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J.M. Rick, B. Kuster, P. Bork, R.B. Russell, G. Superti-Furga (2006)

Proteome survey reveals modularity of the yeast cell machinery. Nature *440(7084)*: 631-636.

Gingras, A.-C., M. Gstaiger, B. Raught, R. Aebersold (2007) Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol *8(8)*: 645-654.

Glaser, F., T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. Bioinformatics *19(1)*: 163-164.

Glickman, M.H., A. Ciechanover (2002) The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction.

Goel, A., D. Colcher, J.-S. Koo, B.J.M. Booth, G. Pavlinkova, S.K. Batra (2000) Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. Biochimica et Biophysica Acta (BBA) - General Subjects *1523(1)*: 13-20.

Gong, L., W.-K. Ji, X.-H. Hu, W.-F. Hu, X.-C. Tang, Z.-X. Huang, L. Li, M. Liu, S.-H. Xiang, E. Wu, Z. Woodward, Y.-Z. Liu, Q.D. Nguyen, D.W.-C. Li (2014) Sumoylation differentially regulates Sp1 to control cell differentiation. Proceedings of the National Academy of Sciences *111(15)*: 5574-5579.

Goshe, M.B., R.D. Smith (2003) Stable isotope-coded proteomic mass spectrometry. Current Opinion in Biotechnology *14(1)*: 101-109.

Govin, J., C. Caron, C. Lestrat, S. Rousseaux, S. Khochbin (2004) The role of histones in chromatin remodelling during mammalian spermiogenesis. European Journal of Biochemistry *271(17)*: 3459-3469.

Griffis, E.R., N. Altan, J. Lippincott-Schwartz, M.A. Powers (2002) Nup98 Is a Mobile Nucleoporin with Transcription-dependent Dynamics. Molecular Biology of the Cell *13(4)*: 1282-1297.

Griffis, E.R., S. Xu, M.A. Powers (2003) Nup98 Localizes to Both Nuclear and Cytoplasmic Sides of the Nuclear Pore and Binds to Two Distinct Nucleoporin Subcomplexes. Molecular Biology of the Cell *14(2)*: 600-610.

Grigoriev, A. (2001) A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. Nucleic Acids Research *29(17)*: 3513-3519.

Gruhler, A., J.V. Olsen, S. Mohammed, P. Mortensen, N.J. Færgeman, M. Mann, O.N. Jensen (2005) Quantitative Phosphoproteomics Applied to the Yeast

Pheromone Signaling Pathway. Molecular & Cellular Proteomics *4(3)*: 310-327.

Guerrero, C., C. Tagwerker, P. Kaiser, L. Huang (2006) An Integrated Mass Spectrometry-based Proteomic Approach: Quantitative Analysis of Tandem Affinity-purified in vivo Cross-linked Protein Complexes (qtax) to Decipher the 26 s Proteasome-interacting Network. Molecular & Cellular Proteomics *5(2)*: 366-378.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotech *17(10)*: 994-999.

Gygi, S.P., Y. Rochon, B.R. Franza, R. Aebersold (1999b) Correlation between Protein and mRNA Abundance in Yeast. Mol Cell Biol *19(3)*: 1720-1730.

Haas, W., B.K. Faherty, S.A. Gerber, J.E. Elias, S.A. Beausoleil, C.E. Bakalarski, X. Li, J. Villén, S.P. Gygi (2006) Optimization and Use of Peptide Mass Measurement Accuracy in Shotgun Proteomics. Molecular & Cellular Proteomics *5(7)*: 1326-1337.

Haglund, K., I. Dikic (2005) Ubiquitylation and cell signaling.

Hanke, S., H. Besir, D. Oesterhelt, M. Mann (2008) Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level. Journal of Proteome Research *7(3)*: 1118-1130.

Harel, A., A.V. Orjalo, T. Vincent, A. Lachish-Zalait, S. Vasu, S. Shah, E. Zimmerman, M. Elbaum, D.J. Forbes (2003) Removal of a Single Pore Subcomplex Results in Vertebrate Nuclei Devoid of Nuclear Pores. Molecular cell *11(4)*: 853-864.

Haupt, Y., R. Maya, A. Kazaz, M. Oren (1997) Mdm2 promotes the rapid degradation of p53. Nature *387(6630)*: 296-299.

Havugimana, Pierre C., G.T. Hart, T. Nepusz, H. Yang, Andrei L. Turinsky, Z. Li, Peggy I. Wang, Daniel R. Boutz, V. Fong, S. Phanse, M. Babu, Stephanie A. Craig, P. Hu, C. Wan, J. Vlasblom, V.-u.-N. Dar, A. Bezginov, Gregory W. Clark, Gabriel C. Wu, Shoshana J. Wodak, Elisabeth R.M. Tillier, A. Paccanaro, Edward M. Marcotte, A. Emili (2012) A Census of Human Soluble Protein Complexes. Cell *150(5)*: 1068-1081.

Hayakawa, T., Y. Ohtani, N. Hayakawa, K. Shinmyozu, M. Saito, F. Ishikawa, J.-i. Nakayama (2007) RBP2 is an MRG15 complex component and down-

regulates intragenic histone H3 lysine 4 methylation. Genes to Cells *12(6)*: 811-826.

Hayakawa, T., F. Zhang, N. Hayakawa, Y. Ohtani, K. Shinmyozu, J.-i. Nakayama, P.R. Andreassen (2010) MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. Journal of Cell Science *123(7)*: 1124-1130.

Hayashihara, K., S. Uchiyama, S. Shimamoto, S. Kobayashi, M. Tomschik, H. Wakamatsu, D. No, H. Sugahara, N. Hori, M. Noda, T. Ohkubo, J. Zlatanova, S. Matsunaga, K. Fukui (2010) The Middle Region of an HP1-binding Protein, HP1-BP74, Associates with Linker DNA at the Entry/Exit Site of Nucleosomal DNA. Journal of Biological Chemistry *285(9)*: 6498-6507.

Heide, H., L. Bleier, M. Steger, J. Ackermann, S. Dröse, B. Schwamb, M. Zörnig, Andreas S. Reichert, I. Koch, I. Wittig, U. Brandt (2012) Complexome Profiling Identifies TMEM126B as a Component of the Mitochondrial Complex I Assembly Complex. Cell Metabolism *16(4)*: 538-549.

Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W.V. Hogue, D. Figeys, M. Tyers (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415(6868)*: 180-183.

Hodel, A.E., M.R. Hodel, E.R. Griffis, K.A. Hennig, G.A. Ratner, S. Xu, M.A. Powers (2002) The Three-Dimensional Structure of the Autoproteolytic, Nuclear Pore-Targeting Domain of the Human Nucleoporin Nup98. Molecular cell *10(2)*: 347-358.

Hornbeck, P.V., J.M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, M. Sullivan (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Research *40(D1)*: D261-D270.

Hosur, R., J. Xu, J. Bienkowska, B. Berger (2011) iWRAP: An Interface Threading Approach with Application to Prediction of Cancer-Related Protein–Protein Interactions. Journal of Molecular Biology *405(5)*: 1295-1310.

Hu, Q., R.J. Noll, H. Li, A. Makarov, M. Hardman, R. Graham Cooks (2005) The Orbitrap: a new mass spectrometer. Journal of Mass Spectrometry *40(4)*: 430-443.

Huang, D.W., B.T. Sherman, R.A. Lempicki (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols *4(1)*: 44-57.

Hubner, N.C., A.W. Bird, J. Cox, B. Splettstoesser, P. Bandilla, I. Poser, A. Hyman, M. Mann (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. The Journal of Cell Biology *189(4)*: 739-754.

Humphries, J.D., A. Byron, M.D. Bass, S.E. Craig, J.W. Pinney, D. Knight, M.J. Humphries (2009) Proteomic analysis of integrin-associated complexes identifies RCC2 as a dual regulator of Rac1 and Arf6. Sci Signal *2(87)*: ra51.

Humphries, J.D., A. Byron, M.J. Humphries (2006) Integrin ligands at a glance. Journal of Cell Science *119(19)*: 3901-3903.

Hurst, G., T. Lankford, S. Kennel (2004) Mass spectrometric detection of affinity purified crosslinked presented. J Am Soc Mass Spectrom *15(6)*: 832-839.

Hynes, R.O. (1987) Integrins: A family of cell surface receptors. Cell *48(4)*: 549-554.

Ibarrola, N., H. Molina, A. Iwahori, A. Pandey (2004) A Novel Proteomic Approach for Specific Identification of Tyrosine Kinase Substrates Using [13C]Tyrosine. Journal of Biological Chemistry *279(16)*: 15805-15813.

Ishibashi, T., H. Hayakawa, R. Ito, M. Miyazawa, Y. Yamagata, M. Sekiguchi (2005) Mammalian enzymes for preventing transcriptional errors caused by oxidative damage. Nucleic Acids Research *33(12)*: 3779-3784.

Jacquemet, G., M.R. Morgan, A. Byron, J.D. Humphries, C.K. Choi, C.S. Chen, P.T. Caswell, M.J. Humphries (2013) Rac1 is deactivated at integrin activation sites through an IQGAP1–filamin-A–RacGAP1 pathway. Journal of Cell Science *126(18)*: 4121-4135.

Jelinic, P., J. Pellegrino, G. David (2011) A Novel Mammalian Complex Containing Sin3B Mitigates Histone Acetylation and RNA Polymerase II Progression within Transcribed Loci. Mol Cell Biol *31(1)*: 54-62.

Jiang, H.-Y., R.C. Wek (2005) Phosphorylation of the α-Subunit of the Eukaryotic Initiation Factor-2 (eIF2α) Reduces Protein Synthesis and Enhances Apoptosis in Response to Proteasome Inhibition. Journal of Biological Chemistry *280(14)*: 14189-14202.

Jiang, H.Y., R.J. Hickey, W. Abdel-Aziz, T.D. Tom, P.W. Wills, J. Liu, L.H. Malkas (2002) Human cell DNA replication is mediated by a discrete multiprotein complex. Journal of Cellular Biochemistry *85(4)*: 762-774.

Johnson, L.N. (2009) The regulation of protein phosphorylation. Biochem Soc Trans *37(Pt 4)*: 627-641.

Kalisman, N., C.M. Adams, M. Levitt (2012) Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. Proceedings of the National Academy of Sciences *109(8)*: 2884-2889.

Kaur, M., M.M. Khan, A. Kar, A. Sharma, S. Saxena (2012) CRL4–DDB1–VPRBP ubiquitin ligase mediates the stress triggered proteolysis of Mcm10. Nucleic Acids Research *40(15)*: 7332-7346.

Kelleher, N.L. (2004) Top-down proteomics. Analytical Chemistry *76(11)*: 197A-203A.

Kelley, L.A., M.J.E. Sternberg (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protocols *4(3)*: 363-371.

Khan, A.A., T. Hanada, M. Mohseni, J.-J. Jeong, L. Zeng, M. Gaetani, D. Li, B.C. Reed, D.W. Speicher, A.H. Chishti (2008) Dematin and Adducin Provide a Novel Link between the Spectrin Cytoskeleton and Human Erythrocyte Membrane by Directly Interacting with Glucose Transporter-1. Journal of Biological Chemistry *283(21)*: 14600-14609.

Khoury, G.A., R.C. Baliban, C.A. Floudas (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep *1*.

Kim, D.-H., D.D. Sarbassov, S.M. Ali, J.E. King, R.R. Latek, H. Erdjument-Bromage, P. Tempst, D.M. Sabatini (2002) mTOR Interacts with Raptor to Form a Nutrient-Sensitive Complex that Signals to the Cell Growth Machinery. Cell *110(2)*: 163-175.

Kim, S., P. Wong, P.A. Coulombe (2006) A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. Nature *441(7091)*: 362-365.

Kirkin, V., T. Lamark, Y.-S. Sou, G. Bjørkøy, J.L. Nunn, J.-A. Bruun, E. Shvets, D.G. McEwan, T.H. Clausen, P. Wild, I. Bilusic, J.-P. Theurillat, A. Øvervatn, T. Ishii, Z. Elazar, M. Komatsu, I. Dikic, T. Johansen (2009) A Role for NBR1 in Autophagosomal Degradation of Ubiquitinated Substrates. Molecular cell *33(4)*: 505-516.

Kirkwood, K.J., Y. Ahmad, M. Larance, A.I. Lamond (2013) Characterization of Native Protein Complexes and Protein Isoform Variation Using Size-fractionation-based Quantitative Proteomics. Molecular & Cellular Proteomics *12(12)*: 3851-3873.

Kouzarides, T. (2007) Chromatin Modifications and Their Function. Cell *128(4)*: 693-705.

Kristensen, A.R., J. Gsponer, L.J. Foster (2012) A high-throughput approach for measuring temporal changes in the interactome. Nat Meth *9(9)*: 907-909.

Krüger, M., M. Moser, S. Ussar, I. Thievessen, C.A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fässler, M. Mann SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. Cell *134(2)*: 353-364.

Kuma, A., M. Hatano, M. Matsui, A. Yamamoto, H. Nakaya, T. Yoshimori, Y. Ohsumi, T. Tokuhisa, N. Mizushima (2004) The role of autophagy during the early neonatal starvation period. Nature *432(7020)*: 1032-1036.

Kumar, G.S., T. Xie, Y. Zhang, I. Radhakrishnan (2011) Solution Structure of the mSin3A PAH2–Pf1 SID1 Complex: A Mad1/Mxd1-Like Interaction Disrupted by MRG15 in the Rpd3S/Sin3S Complex. Journal of Molecular Biology *408(5)*: 987-1000.

Kunjappu, M.J., M. Hochstrasser (2014) Assembly of the 20S proteasome. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1843(1)*: 2-12.

Kuo, M.-H., C.D. Allis (1998) Roles of histone acetyltransferases and deacetylases in gene regulation. BioEssays *20(8)*: 615-626.

Kuo, M.-H., C.D. Allis (1999) In Vivo Cross-Linking and Immunoprecipitation for Studying Dynamic Protein:DNA Associations in a Chromatin Environment. Methods *19(3)*: 425-433.

Lamb, C.A., T. Yoshimori, S.A. Tooze (2013) The autophagosome: origins unknown, biogenesis complex. Nat Rev Mol Cell Biol *14(12)*: 759-774.

Lander, E.S. (2011) Initial impact of the sequencing of the human genome. Nature *470(7333)*: 187-197.

Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E.J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson, G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, M.V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G.A. Evans, M. Athanasiou, R. Schultz, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K.

Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M. Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.P. Yang, R.F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, A. Patrinos, M.J. Morgan, P. de Jong, J.J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y.J. Chen (2001) Initial sequencing and analysis of the human genome. Nature *409(6822)*: 860-921.

Lappe, M., L. Holm (2004) Unraveling protein interaction networks with near-optimal efficiency. Nat Biotech *22(1)*: 98-103.

Larance, M., Y. Ahmad, K.J. Kirkwood, T. Ly, A.I. Lamond (2013) Global Subcellular Characterization of Protein Degradation Using Quantitative Proteomics. Molecular & Cellular Proteomics *12(3)*: 638-650.

Larance, M., A.P. Bailly, E. Pourkarimi, R.T. Hay, G. Buchanan, S. Coulthurst, D.P. Xirodimas, A. Gartner, A.I. Lamond (2011) Stable-isotope labeling with amino acids in nematodes. Nat Meth *8(10)*: 849-851.

Larance, M., K.J. Kirkwood, D.P. Xirodimas, E. Lundberg, M. Uhlen, A.I. Lamond (2012) Characterization of MRFAP1 Turnover and Interactions Downstream of the NEDD8 Pathway. Molecular & Cellular Proteomics *11(3)*.

Larivière, L., C. Plaschka, M. Seizl, E.V. Petrotchenko, L. Wenzeck, C.H. Borchers, P. Cramer (2013) Model of the Mediator middle module based on protein cross-linking. Nucleic Acids Research *41(20)*: 9266-9273.

LaSarge, C.L., S.C. Danzer (2014) Mechanisms regulating neuronal excitability and seizure development following mTOR pathway hyperactivation. Frontiers in Molecular Neuroscience *7*.

Lasker, K., F. Förster, S. Bohn, T. Walzthoeni, E. Villa, P. Unverdorben, F. Beck, R. Aebersold, A. Sali, W. Baumeister (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. Proceedings of the National Academy of Sciences *109(5)*: 1380-1387.

Lauber, M.A., J. Rappsilber, J.P. Reilly (2012) Dynamics of Ribosomal Protein S1 on a Bacterial Ribosome with Cross-Linking and Mass Spectrometry. Molecular & Cellular Proteomics *11(12)*: 1965-1976.

Le Rouzic, E., N. Belaïdouni, E. Estrabaud, M. Morel, J.-C. Rain, C. Transy, F. Margottin-Goguet (2007) HIV1 Vpr Arrests the Cell Cycle by Recruiting DCAF1/VprBP, a Receptor of the Cul4-DDB1 Ubiquitin Ligase. Cell Cycle *6(2)*: 182-188.

Lee, J.-S., A. Shilatifard (2007) A site to remember: H3K36 methylation a mark for histone deacetylation. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *618(1–2)*: 130-134.

Leidecker, O., I. Matic, B. Mahata, E. Pion, D.P. Xirodimas (2012) The ubiquitin E1 enzyme Ube1 mediates NEDD8 activation under diverse stress conditions. Cell Cycle *11(6)*: 1142-1150.

Leitner, A., Lukasz A. Joachimiak, A. Bracher, L. Mönkemeyer, T. Walzthoeni, B. Chen, S. Pechmann, S. Holmes, Y. Cong, B. Ma, S. Ludtke, W. Chiu, F.U. Hartl, R. Aebersold, J. Frydman (2012) The Molecular Architecture of the Eukaryotic Chaperonin TRiC/CCT. Structure (London, England : 1993) *20(5)*: 814-825.

Leung, J.K., N. Berube, S. Venable, S. Ahmed, N. Timchenko, O.M. Pereira-Smith (2001) MRG15 Activates the B-myb Promoter through Formation of a Nuclear Complex with the Retinoblastoma Protein and the Novel Protein PAM14. Journal of Biological Chemistry *276(42)*: 39171-39178.

Liao, H., X.J. Liu, J.L. Blank, D.C. Bouck, H. Bernard, K. Garcia, E.S. Lightcap (2011) Quantitative Proteomic Analysis of Cellular Protein Modulation upon Inhibition of the NEDD8-Activating Enzyme by MLN4924. Molecular & Cellular Proteomics *10(11)*.

Lichtarge, O., M.E. Sowa (2002) Evolutionary predictions of binding surfaces and interactions. Current Opinion in Structural Biology *12(1)*: 21-27.

Link, A.J., J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates (1999) Direct analysis of protein complexes using mass spectrometry. Nat Biotech *17(7)*: 676-682.

Loïodice, I., A. Alves, G. Rabut, M. van Overbeek, J. Ellenberg, J.-B. Sibarita, V. Doye (2004) The Entire Nup107-160 Complex, Including Three New Members, Is Targeted as One Entity to Kinetochores in Mitosis. Molecular Biology of the Cell *15(7)*: 3333-3344.

Luco, R.F., Q. Pan, K. Tominaga, B.J. Blencowe, O.M. Pereira-Smith, T. Misteli (2010) Regulation of alternative splicing by histone modifications. Science *327(5968)*: 996-1000.

Lundberg, E., L. Fagerberg, D. Klevebring, I. Matic, T. Geiger, J. Cox, C. Algenas, J. Lundeberg, M. Mann, M. Uhlen (2010) Defining the transcriptome and proteome in three functionally different human cell lines. Mol Syst Biol *6*.

Lutzmann, M., R. Kunze, A. Buerer, U. Aebi, E. Hurt (2002) Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. EMBO J *21(3)*: 387-397.

Ly, T., Y. Ahmad, A. Shlien, D. Soroka, A. Mills, M.J. Emanuele, M.R. Stratton, A.I. Lamond (2014) A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. Elife *3*: e01630.

MacBeath, G., S.L. Schreiber (2000) Printing Proteins as Microarrays for High-Throughput Function Determination. Science *289(5485)*: 1760-1763.

Maier, T., M. Güell, L. Serrano (2009) Correlation of mRNA and protein in complex biological samples. FEBS Letters *583(24)*: 3966-3973.

Maiolica, A., D. Cittaro, D. Borsotti, L. Sennels, C. Ciferri, C. Tarricone, A. Musacchio, J. Rappsilber (2007) Structural Analysis of Multiprotein Complexes by Cross-linking, Mass Spectrometry, and Database Searching. Molecular & Cellular Proteomics *6(12)*: 2200-2211.

Malovannaya, A., Rainer B. Lanz, Sung Y. Jung, Y. Bulynko, Nguyen T. Le, Doug W. Chan, C. Ding, Y. Shi, N. Yucer, G. Krenciute, B.-J. Kim, C. Li, R. Chen, W. Li, Y. Wang, Bert W. O'Malley, J. Qin (2011) Analysis of the Human Endogenous Coregulator Complexome. Cell *145(5)*: 787-799.

Malovannaya, A., Y. Li, Y. Bulynko, S.Y. Jung, Y. Wang, R.B. Lanz, B.W. O'Malley, J. Qin (2010) Streamlined analysis schema for high-throughput identification of endogenous protein complexes. Proceedings of the National Academy of Sciences *107(6)*: 2431-2436.

Man, O., Y. Pilpel (2007) Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. Nat Genet *39(3)*: 415-421.

Martinović, S., T.D. Veenstra, G.A. Anderson, L. Paša-Tolić, R.D. Smith (2002) Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. Journal of Mass Spectrometry *37(1)*: 99-107.

Memisevic, V., A. Wallqvist, J. Reifman (2013) Reconstituting protein interaction networks using parameter-dependent domain-domain interactions. BMC Bioinformatics *14*: 154.

Metz, B., G.F.A. Kersten, G.J.E. Baart, A. de Jong, H. Meiring, J. ten Hove, M.J. van Steenbergen, W.E. Hennink, D.J.A. Crommelin, W. Jiskoot (2006) Identification of Formaldehyde-Induced Modifications in Proteins: Reactions with Insulin. Bioconjugate Chemistry *17(3)*: 815-822.

Metz, B., G.F.A. Kersten, P. Hoogerhout, H.F. Brugghe, H.A.M. Timmermans, A. de Jong, H. Meiring, J.t. Hove, W.E. Hennink, D.J.A. Crommelin, W. Jiskoot (2004) Identification of Formaldehyde-induced Modifications in Proteins: REACTIONS WITH MODEL PEPTIDES. Journal of Biological Chemistry *279(8)*: 6235-6243.

Meunier, L., Y.-K. Usherwood, K.T. Chung, L.M. Hendershot (2002) A Subset of Chaperones and Folding Enzymes Form Multiprotein Complexes in Endoplasmic Reticulum to Bind Nascent Proteins. Molecular Biology of the Cell *13(12)*: 4456-4469.

Milhollen, M.A., U. Narayanan, T.A. Soucy, P.O. Veiby, P.G. Smith, B. Amidon (2011) Inhibition of NEDD8-Activating Enzyme Induces Rereplication and Apoptosis in Human Tumor Cells Consistent with Deregulating CDT1 Turnover. Cancer Research *71(8)*: 3042-3051.

Mitchell, R.T., M. E Camacho-Moll, J. Macdonald, R.A. Anderson, C.J.H. Kelnar, M. O'Donnell, R.M. Sharpe, L.B. Smith, K.M. Grigor, W.H.B. Wallace, H. Stoop, K.P. Wolffenbuttel, R. Donat, P.T.K. Saunders, L.H.J. Looijenga (2014) Intratubular germ cell neoplasia of the human testis: heterogeneous protein expression and relation to invasive potential. Mod Pathol.

Morgan, Mark R., H. Hamidi, Mark D. Bass, S. Warwood, C. Ballestrem, Martin J. Humphries (2013) Syndecan-4 Phosphorylation Is a Control Point for Integrin Recycling. Developmental cell *24(5)*: 472-485.

Müller, D.R., P. Schindler, H. Towbin, U. Wirth, H. Voshol, S. Hoving, M.O. Steinmetz (2001) Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. Analytical Chemistry *73(9)*: 1927-1934.

Murakami, K., H. Elmlund, N. Kalisman, D.A. Bushnell, C.M. Adams, M. Azubel, D. Elmlund, Y. Levi-Kalisman, X. Liu, B.J. Gibbons, M. Levitt, R.D. Kornberg

(2013) Architecture of an RNA polymerase II transcription pre-initiation complex. Science *342(6159)*: 1238724.

Murthy, A., Y. Li, I. Peng, M. Reichelt, A.K. Katakam, R. Noubade, M. Roose-Girma, J. DeVoss, L. Diehl, R.R. Graham, M. van Lookeren Campagne (2014) A Crohn/'s disease variant in Atg16l1 enhances its degradation by caspase 3. Nature *506(7489)*: 456-462.

Nagaraj, N., J.R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Paabo, M. Mann (2011) Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol *7*.

Nguyen, Vu Q., A. Ranjan, F. Stengel, D. Wei, R. Aebersold, C. Wu, Andres E. Leschziner (2013) Molecular Architecture of the ATP-Dependent Chromatin-Remodeling Complex SWR1. Cell *154(6)*: 1220-1231.

Nikolov, D.B., S.K. Burley (1997) RNA polymerase II transcription initiation: A structural view. Proceedings of the National Academy of Sciences *94(1)*: 15-22.

Nilsson, P., L. Paavilainen, K. Larsson, J. Ödling, M. Sundberg, A.-C. Andersson, C. Kampf, A. Persson, C.A.-K. Szigyarto, J. Ottosson, E. Björling, S. Hober, H. Wernérus, K. Wester, F. Pontén, M. Uhlen (2005) Towards a human proteome atlas: High-throughput generation of mono-specific antibodies for tissue profiling. PROTEOMICS *5(17)*: 4327-4337.

Nishi, H., K. Hashimoto, Anna R. Panchenko (2011) Phosphorylation in Protein-Protein Binding: Effect on Stability and Function. Structure *19(12)*: 1807-1815.

Novak, P., W.E. Haskins, M.J. Ayson, R.B. Jacobsen, J.S. Schoeniger, M.D. Leavell, M.M. Young, G.H. Kruppa (2005) Unambiguous Assignment of Intramolecular Chemical Cross-Links in Modified Mammalian Membrane Proteins by Fourier Transform-Tandem Mass Spectrometry. Analytical Chemistry *77(16)*: 5101-5106.

Ofran, Y., B. Rost (2003) Predicted protein–protein interaction sites from local sequence information. FEBS Letters *544(1–3)*: 236-239.

Ohsumi, Y. (2006) Protein turnover. IUBMB Life *58(5-6)*: 363-369.

Okamura, H., J. Aramburu, C. García-Rodríguez, J.P.B. Viola, A. Raghavan, M. Tahiliani, X. Zhang, J. Qin, P.G. Hogan, A. Rao Concerted Dephosphorylation of

the Transcription Factor NFAT1 Induces a Conformational Switch that Regulates Transcriptional Activity. Molecular cell *6(3)*: 539-550.

Olinares, P.D.B., L. Ponnala, K.J. van Wijk (2010) Megadalton Complexes in the Chloroplast Stroma of Arabidopsis thaliana Characterized by Size Exclusion Chromatography, Mass Spectrometry, and Hierarchical Clustering. Molecular & Cellular Proteomics *9(7)*: 1594-1615.

Olsen, J.V., B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. Cell *127(3)*: 635-648.

Olsen, J.V., M. Mann (2013) Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry. Molecular & Cellular Proteomics *12(12)*: 3444-3452.

Olsen, J.V., M. Vermeulen, A. Santamaria, C. Kumar, M.L. Miller, L.J. Jensen, F. Gnad, J. Cox, T.S. Jensen, E.A. Nigg, S. Brunak, M. Mann (2010) Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis.

Ong, S.-E., B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann (2002a) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Molecular & Cellular Proteomics *1(5)*: 376-386.

Ong, S.-E., I. Kratchmarova, M. Mann (2002b) Properties of 13C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). Journal of Proteome Research *2(2)*: 173-181.

Orlando, V. (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends in Biochemical Sciences *25(3)*: 99-104.

Pape, T., H. Meka, S. Chen, G. Vicentini, M. van Heel, S. Onesti (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. EMBO reports *4(11)*: 1079-1083.

Pardo, P.S., J.K. Leung, J.C. Lucchesi, O.M. Pereira-Smith (2002) MRG15, a Novel Chromodomain Protein, Is Present in Two Distinct Multiprotein Complexes Involved in Transcriptional Activation. Journal of Biological Chemistry *277(52)*: 50860-50866.

Perkins, D.N., D.J.C. Pappin, D.M. Creasy, J.S. Cottrell (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. ELECTROPHORESIS *20(18)*: 3551-3567.

Peters, J.-M., Z. Cejka, J.R. Harris, J.A. Kleinschmidt, W. Baumeister (1993) Structural Features of the 26 S Proteasome Complex. Journal of Molecular Biology *234(4)*: 932-937.

Petrotchenko, E.V., C.H. Borchers (2010) Crosslinking combined with mass spectrometry for structural proteomics. Mass Spectrometry Reviews *29(6)*: 862-876.

Petrotchenko, E.V., J.J. Serpa, C.H. Borchers (2011) An Isotopically Coded CID-cleavable Biotinylated Cross-linker for Structural Proteomics. Molecular & Cellular Proteomics *10(2)*.

Pickart, C.M., M.J. Eddins (2004) Ubiquitin: structures, functions, mechanisms. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1695(1–3)*: 55-72.

Pratt, J.M., D.M. Simpson, M.K. Doherty, J. Rivers, S.J. Gaskell, R.J. Beynon (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protocols *1(2)*: 1029-1043.

Pritchard, C.E.J., M. Fornerod, L.H. Kasper, J.M.A. van Deursen (1999) RAE1 Is a Shuttling mRNA Export Factor That Binds to a GLEBS-like NUP98 Motif at the Nuclear Pore Complex through Multiple Domains. The Journal of Cell Biology *145(2)*: 237-254.

Rabut, G., M. Peter (2008) Function and regulation of protein neddylation. EMBO Rep *9(10)*: 969-976.

Radu, A., M.S. Moore, G. Blobel (1995) The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. Cell *81(2)*: 215-222.

Rao, V.S., K. Srinivas, G.N. Sujini, G.N. Kumar (2014) Protein-Protein Interaction Detection: Methods and Analysis. Int J Proteomics *2014*: 147648.

Rappsilber, J., S. Siniossoglou, E.C. Hurt, M. Mann (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. Analytical Chemistry *72(2)*: 267-275.

Ren, Y., H.-S. Seo, G. Blobel, A. Hoelz (2010) Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1. Proceedings of the National Academy of Sciences *107(23)*: 10406-10411.

Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotech *17(10)*: 1030-1032.

Rinner, O., J. Seebacher, T. Walzthoeni, L. Mueller, M. Beck, A. Schmidt, M. Mueller, R. Aebersold (2008) Identification of cross-linked peptides from large sequence databases. Nat Meth *5(4)*: 315-318.

Ritorto, M.S., K. Cook, K. Tyagi, P.G.A. Pedrioli, M. Trost (2013) Hydrophilic Strong Anion Exchange (hSAX) Chromatography for Highly Orthogonal Peptide Separation of Complex Proteomes. Journal of Proteome Research *12(6)*: 2449-2457.

Rivera, C., Z.A. Gurard-Levin, G. Almouzni, A. Loyola (2014) Histone lysine methylation and chromatin replication. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*(0)*.

Robinson, P.J.J., D.A. Bushnell, M.J. Trnka, A.L. Burlingame, R.D. Kornberg (2012) Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II. Proceedings of the National Academy of Sciences *109(44)*: 17931-17935.

Ross, P.L., Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. Molecular & Cellular Proteomics *3(12)*: 1154-1169.

Ruepp, A., B. Waegele, M. Lechner, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, H.-W. Mewes (2010) CORUM: the comprehensive resource of mammalian protein complexes—2009. Nucleic Acids Research *38(suppl 1)*: D497-D501.

Rumlová, M., J. Benedíková, R. Cubínková, I. Pichová, T. Ruml (2001) Comparison of Classical and Affinity Purification Techniques of Mason-Pfizer Monkey Virus Capsid Protein: The Alteration of the Product by an Affinity Tag. Protein Expression and Purification *23(1)*: 75-83.

Salwinski, L., D. Eisenberg (2003) Computational methods of analysis of protein–protein interactions. Current Opinion in Structural Biology *13(3)*: 377-382.

Schmitt-Ulms, G., K. Hansen, J. Liu, C. Cowdrey, J. Yang, S.J. DeArmond, F.E. Cohen, S.B. Prusiner, M.A. Baldwin (2004) Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. Nat Biotech *22(6)*: 724-731.

Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach (2011) Global quantification of mammalian gene expression control. Nature *473(7347)*: 337-342.

Sekiguchi, T., R. Ito, H. Hayakawa, M. Sekiguchi (2013) Elimination and Utilization of Oxidized Guanine Nucleotides in the Synthesis of RNA and Its Precursors. Journal of Biological Chemistry *288(12)*: 8128-8135.

Sharon, M., T. Taverner, X.I. Ambroggio, R.J. Deshaies, C.V. Robinson (2006) Structural Organization of the 19S Proteasome Lid: Insights from MS of Intact Complexes. PLoS Biol *4(8)*: e267.

Shattil, S.J. (2005) Integrins and Src: dynamic duo of adhesion signaling. Trends in Cell Biology *15(8)*: 399-403.

Shevchenko, A., M. Wilm, O. Vorm, M. Mann (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. Analytical Chemistry *68(5)*: 850-858.

Sigala, J.L.D., V. Bottero, D.B. Young, A. Shevchenko, F. Mercurio, I.M. Verma (2004) Activation of Transcription Factor NF-κB Requires ELKS, an IκB Kinase Regulatory Subunit. Science *304(5679)*: 1963-1967.

Singh, R., D. Park, J. Xu, R. Hosur, B. Berger (2010) Struct2Net: a web service to predict protein–protein interactions using a structure-based approach. Nucleic Acids Research *38(suppl 2)*: W508-W515.

Siniossoglou, S., M. Lutzmann, H. Santos-Rosa, K. Leonard, S. Mueller, U. Aebi, E. Hurt (2000) Structure and Assembly of the Nup84p Complex. The Journal of Cell Biology *149(1)*: 41-54.

Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. Mass Spectrometry Reviews *25(4)*: 663-682.

Snel, B., G. Lehmann, P. Bork, M.A. Huynen (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Research *28(18)*: 3442-3444.

Soucy, T.A., P.G. Smith, M.A. Milhollen, A.J. Berger, J.M. Gavin, S. Adhikari, J.E. Brownell, K.E. Burke, D.P. Cardin, S. Critchley, C.A. Cullis, A. Doucette, J.J. Garnsey, J.L. Gaulin, R.E. Gershman, A.R. Lublinsky, A. McDonald, H. Mizutani, U. Narayanan, E.J. Olhava, S. Peluso, M. Rezaei, M.D. Sintchak, T. Talreja, M.P. Thomas, T. Traore, S. Vyskocil, G.S. Weatherhead, J. Yu, J. Zhang, L.R. Dick, C.F. Claiborne, M. Rolfe, J.B. Bolen, S.P. Langston (2009) An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. Nature *458(7239)*: 732-736.

Spruill, L.S., P.J. McDermott (2009) Role of the 5′-untranslated region in regulating translational efficiency of specific mRNAs in adult cardiocytes. The FASEB Journal *23(9)*: 2879-2887.

Stelzl, U., E.E. Wanker (2006) The value of high quality protein–protein interaction networks for systems biology. Current Opinion in Chemical Biology *10(6)*: 551-558.

Suntharalingam, M., S.R. Wente (2003) Peering through the Pore: Nuclear Pore Complex Structure, Assembly, and Function. Developmental cell *4(6)*: 775-789.

Sutherland, B.W., J. Toews, J. Kast (2008) Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions. Journal of Mass Spectrometry *43(6)*: 699-715.

Swords, R.T., K.R. Kelly, P.G. Smith, J.J. Garnsey, D. Mahalingam, E. Medina, K. Oberheu, S. Padmanabhan, M. O'Dwyer, S.T. Nawrocki, F.J. Giles, J.S. Carew (2010) Inhibition of NEDD8-activating enzyme: a novel approach for the treatment of acute myeloid leukemia. Blood *115(18)*: 3796-3800.

Sy, S.M.-H., M.S.Y. Huen, J. Chen (2009) MRG15 Is a Novel PALB2-interacting Factor Involved in Homologous Recombination. Journal of Biological Chemistry *284(32)*: 21127-21131.

Szklarczyk, D., A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L.J. Jensen, C.v. Mering (2010) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Research.

Tagwerker, C., K. Flick, M. Cui, C. Guerrero, Y. Dou, B. Auer, P. Baldi, L. Huang, P. Kaiser (2006) A Tandem Affinity Tag for Two-step Purification under Fully Denaturing Conditions: Application in Ubiquitin Profiling and Protein Complex Identification Combined with in vivoCross-Linking. Molecular & Cellular Proteomics *5(4)*: 737-748.

Tammsalu, T., I. Matic, E.G. Jaffray, A.F.M. Ibrahim, M.H. Tatham, R.T. Hay (2014) Proteome-Wide Identification of SUMO2 Modification Sites.

Tao, W.A., R. Aebersold (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. Current Opinion in Biotechnology *14(1)*: 110-118.

Tatham, M.H., I. Matic, M. Mann, R.T. Hay (2011) Comparative Proteomic Analysis Identifies a Role for SUMO in Protein Quality Control.

Thompson, J.D., D.G. Higgins, T.J. Gibson (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research *22(22)*: 4673-4680.

Tikhmyanova, N., E. Izumchenko, I. Serebriiskii, E. Golemis (2007) A Bacterial/Yeast Merged Two-Hybrid System; in Ochs, M. (ed): Gene Function Analysis, Humana Press, pp 257-290.

Toews, J., J.C. Rogalski, T.J. Clark, J. Kast (2008) Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions. Analytica Chimica Acta *618(2)*: 168-183.

Tokunaga, F., S.-i. Sakata, Y. Saeki, Y. Satomi, T. Kirisako, K. Kamei, T. Nakagawa, M. Kato, S. Murata, S. Yamaoka, M. Yamamoto, S. Akira, T. Takao, K. Tanaka, K. Iwai (2009) Involvement of linear polyubiquitylation of NEMO in NF-[kappa]B activation. Nat Cell Biol *11(2)*: 123-132.

Tominaga, K., B. Kirtane, J.G. Jackson, Y. Ikeno, T. Ikeda, C. Hawks, J.R. Smith, M.M. Matzuk, O.M. Pereira-Smith (2005) MRG15 Regulates Embryonic Development and Cell Proliferation. Mol Cell Biol *25(8)*: 2924-2937.

Tominaga, K., D.M. Magee, M.M. Matzuk, O.M. Pereira-Smith (2004) PAM14, a Novel MRG- and Rb-Associated Protein, Is Not Required for Development and T-Cell Function in Mice. Mol Cell Biol *24(19)*: 8366-8373.

Tootle, T.L., I. Rebay (2005) Post-translational modifications influence transcription factor activity: A view from the ETS superfamily. BioEssays *27(3)*: 285-298.

Tosi, A., C. Haas, F. Herzog, A. Gilmozzi, O. Berninghausen, C. Ungewickell, Christian B. Gerhold, K. Lakomek, R. Aebersold, R. Beckmann, K.-P. Hopfner (2013) Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex. Cell *154(6)*: 1207-1219.

Toyama, Brandon H., Jeffrey N. Savas, Sung K. Park, Michael S. Harris, Nicholas T. Ingolia, John R. Yates, III, Martin W. Hetzer Identification of Long-Lived Proteins Reveals Exceptional Stability of Essential Cellular Structures. Cell *154(5)*: 971-982.

Trinkle-Mulcahy, L., J. Andersen, Y.W. Lam, G. Moorhead, M. Mann, A.I. Lamond (2006) Repo-Man recruits PP1γ to chromatin and is essential for cell viability. The Journal of Cell Biology *172(5)*: 679-692.

Trinkle-Mulcahy, L., S. Boulon, Y.W. Lam, R. Urcia, F.-M. Boisvert, F. Vandermoere, N.A. Morrice, S. Swift, U. Rothbauer, H. Leonhardt, A. Lamond (2008) Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. The Journal of Cell Biology *183(2)*: 223-239.

Uhlén, M., E. Björling, C. Agaton, C.A.-K. Szigyarto, B. Amini, E. Andersen, A.-C. Andersson, P. Angelidou, A. Asplund, C. Asplund, L. Berglund, K. Bergström, H. Brumer, D. Cerjan, M. Ekström, A. Elobeid, C. Eriksson, L. Fagerberg, R. Falk, J. Fall, M. Forsberg, M.G. Björklund, K. Gumbel, A. Halimi, I. Hallin, C. Hamsten, M. Hansson, M. Hedhammar, G. Hercules, C. Kampf, K. Larsson, M. Lindskog, W. Lodewyckx, J. Lund, J. Lundeberg, K. Magnusson, E. Malm, P. Nilsson, J. Ödling, P. Oksvold, I. Olsson, E. Öster, J. Ottosson, L. Paavilainen, A. Persson, R. Rimini, J. Rockberg, M. Runeson, Å. Sivertsson, A. Sköllermo, J. Steen, M. Stenvall, F. Sterky, S. Strömberg, M. Sundberg, H. Tegel, S. Tourle, E. Wahlund, A. Waldén, J. Wan, H. Wernérus, J. Westberg, K. Wester, U. Wrethagen, L.L. Xu, S. Hober, F. Pontén (2005) A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. Molecular & Cellular Proteomics *4(12)*: 1920-1932.

Ungermann, C., D. Langosch (2005) Functions of SNAREs in intracellular membrane fusion and lipid bilayer mixing. Journal of Cell Science *118(17)*: 3819-3828.

Valpuesta, J.M., J. Martín-Benito, P. Gómez-Puertas, J.L. Carrascosa, K.R. Willison (2002) Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. FEBS Letters *529(1)*: 11-16.

Varshavsky, A. (1996) The N-end rule: functions, mysteries, uses. Proceedings of the National Academy of Sciences *93(22)*: 12142-12149.

Vasilescu, J., X. Guo, J. Kast (2004) Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. PROTEOMICS *4(12)*: 3845-3854.

Vasu, S., S. Shah, A. Orjalo, M. Park, W.H. Fischer, D.J. Forbes (2001) Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. The Journal of Cell Biology *155(3)*: 339-354.

Verschoor, A., S. Srivastava, R. Grassucci, J. Frank (1996) Native 3D structure of eukaryotic 80s ribosome: morphological homology with E. coli 70S ribosome. The Journal of Cell Biology *133(3)*: 495-505.

Vogel, C., R. de Sousa Abreu, D. Ko, S.Y. Le, B.A. Shapiro, S.C. Burns, D. Sandhu, D.R. Boutz, E.M. Marcotte, L.O. Penalva (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line.

von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, P. Bork (2002) Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417(6887)*: 399-403.

Wagner, S.A., P. Beli, B.T. Weinert, M.L. Nielsen, J. Cox, M. Mann, C. Choudhary (2011) A Proteome-wide, Quantitative Survey of In Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles. Molecular & Cellular Proteomics *10(10)*.

Waterhouse, A.M., J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25(9)*: 1189-1191.

Watson, I.R., M.S. Irwin, M. Ohh (2011) NEDD8 Pathways in Cancer, Sine Quibus Non. Cancer Cell *19(2)*: 168-176.

Wen, X., K.M. Duus, T.D. Friedrich, C.M.C. de Noronha (2007) The HIV1 Protein Vpr Acts to Promote G2 Cell Cycle Arrest by Engaging a DDB1 and Cullin4A-containing Ubiquitin Ligase Complex Using VprBP/DCAF1 as an Adaptor. Journal of Biological Chemistry *282(37)*: 27046-27057.

Wessel, D., U.I. Flügge (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Analytical Biochemistry *138(1)*: 141-143.

Wheaton, R.M., W.C. Bauman (1953) NON-IONIC SEPARATIONS WITH ION EXCHANGE RESINS. Annals of the New York Academy of Sciences *57(3)*: 159-176.

Wilkinson, K.D. (1997) Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. The FASEB Journal *11(14)*: 1245-1256.

Wittig, I., H.-P. Braun, H. Schagger (2006) Blue native PAGE. Nat Protocols *1(1)*: 418-428.

Wojcik, J., V. Schächter (2001) Protein-protein interaction map inference using interacting domain profile pairs. Bioinformatics *17(suppl 1)*: S296-S305.

Wolf-Yadlin, A., M. Sevecka, G. MacBeath (2009) Dissecting protein function and signaling using protein microarrays. Current Opinion in Chemical Biology *13(4)*: 398-405.

Wolfenson, H., I. Lavelin, B. Geiger (2013) Dynamic Regulation of the Structure and Functions of Integrin Adhesions. Developmental cell *24(5)*: 447-458.

Wullschleger, S., R. Loewith, M.N. Hall (2006) TOR Signaling in Growth and Metabolism. Cell *124(3)*: 471-484.

Xie, T., R. Graveline, Ganesan S. Kumar, Y. Zhang, A. Krishnan, G. David, I. Radhakrishnan (2012) Structural Basis for Molecular Interactions Involving MRG Domains: Implications in Chromatin Biology. Structure *20(1)*: 151-160.

Xirodimas, D.P. (2008) Novel substrates and functions for the ubiquitin-like molecule NEDD8. Biochem Soc Trans *36(Pt 5)*: 802-806.

Xu, H., P.-H. Hsu, L. Zhang, M.-D. Tsai, M.A. Freitas (2010) Database Search Algorithm for Identification of Intact Cross-Links in Proteins and Peptides Using Tandem Mass Spectrometry. Journal of Proteome Research *9(7)*: 3384-3393.

Xu, P., H. Tan, D.M. Duong, Y. Yang, J. Kupsco, K.H. Moberg, H. Li, P. Jin, J. Peng (2012) Stable Isotope Labeling with Amino Acids in Drosophila for

Quantifying Proteins and Modifications. Journal of Proteome Research *11(9)*: 4403-4412.

Yang, B., Y.-J. Wu, M. Zhu, S.-B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.-X. Li, H.-F. Chen, S.-K. Luo, Y.-H. Ding, L.-H. Wang, Z. Hao, L.-Y. Xiu, S. Chen, K. Ye, S.-M. He, M.-Q. Dong (2012) Identification of cross-linked peptides from complex samples. Nat Meth *9(9)*: 904-906.

Yap, K.L., M.-M. Zhou (2011) Structure and Mechanisms of Lysine Methylation Recognition by the Chromodomain in Gene Transcription. Biochemistry *50(12)*: 1966-1980.

Ye, F., F. Lagarrigue, M.H. Ginsberg (2014) SnapShot: Talin and the Modular Nature of the Integrin Adhesome. Cell *156(6)*: 1340-1340.e1341.

Yochum, G.S., D.E. Ayer (2002) Role for the Mortality Factors MORF4, MRGX, and MRG15 in Transcriptional Repression via Associations with Pf1, mSin3A, and Transducin-Like Enhancer of Split. Mol Cell Biol *22(22)*: 7868-7876.

Zeiler, M., W.L. Straube, E. Lundberg, M. Uhlen, M. Mann (2012) A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines. Molecular & Cellular Proteomics *11(3)*.

Zha, M., C. Zhong, Y. Peng, H. Hu, J. Ding (2006) Crystal Structures of Human NUDT5 Reveal Insights into the Structural Basis of the Substrate Specificity. Journal of Molecular Biology *364(5)*: 1021-1033.

Zhang, P., J. Du, B. Sun, X. Dong, G. Xu, J. Zhou, Q. Huang, Q. Liu, Q. Hao, J. Ding (2006a) Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Research *34(22)*: 6621-6628.

Zhang, P., J. Zhao, B. Wang, J. Du, Y. Lu, J. Chen, J. Ding (2006b) The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. Protein Science *15(10)*: 2423-2434.

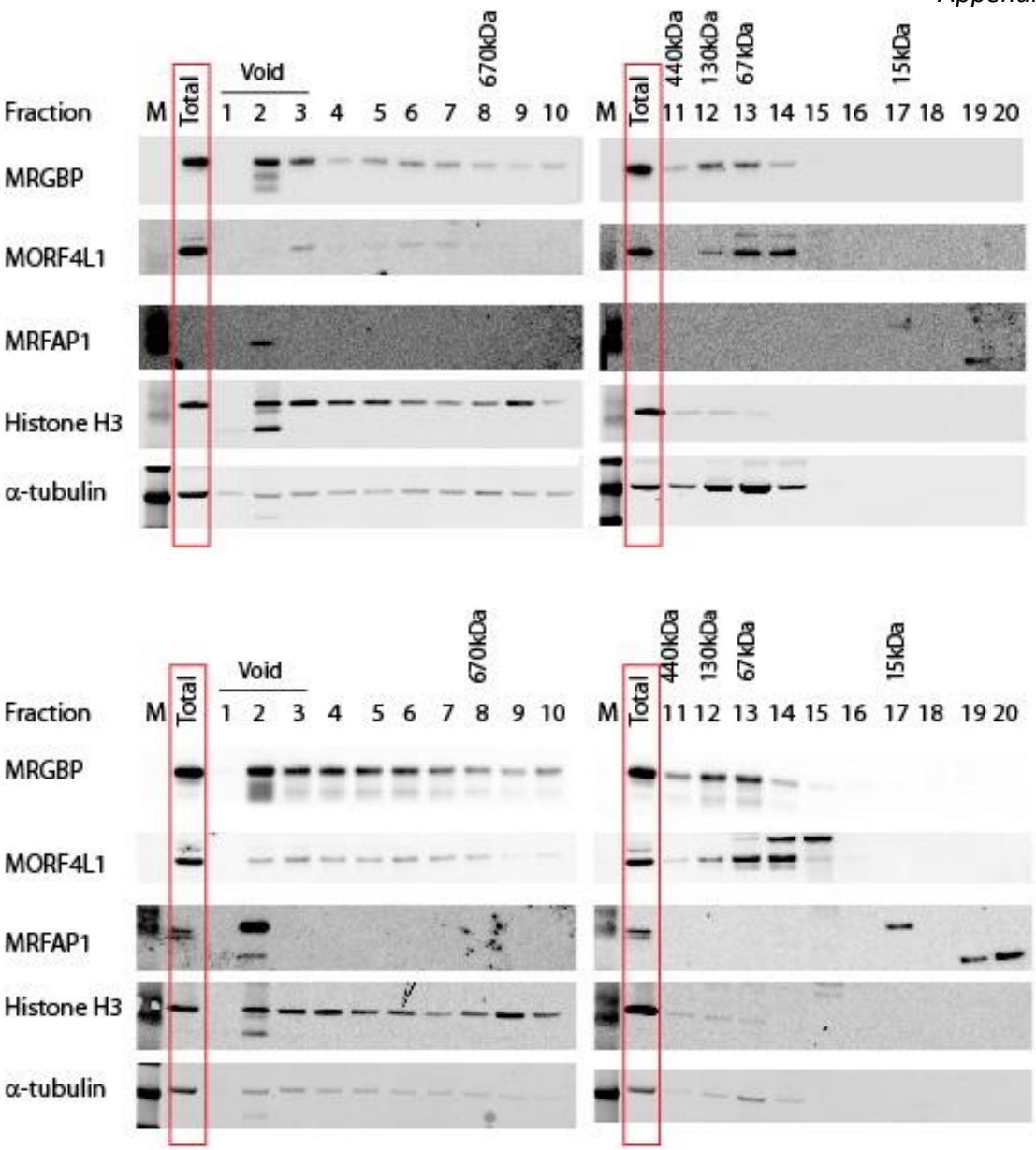
Zhang, X., H. Wen, X. Shi (2012) Lysine methylation: beyond histones. Acta Biochimica et Biophysica Sinica *44(1)*: 14-27.

Zotti, T., I. Scudiero, P. Settembre, A. Ferravante, P. Mazzone, L. D'Andrea, C. Reale, P. Vito, R. Stilo (2014) TRAF6-mediated ubiquitination of NEMO requires p62/sequestosome-1. Molecular Immunology *58(1)*: 27-31.

Appendix I

Native SEC can be used to study protein complexes in testis

To test whether the SEC approach I developed could be utilised in whole tissues, I lysed whole testes from mice aged 48 days and 160 days in PBS and prepared them for SEC as previously for U2OS cell lysates (see section 2.7.3). The lysate chromatogram was strikingly similar to those in U2OS cells indicating that the range of protein complex sizes is similar in tissues to the cell culture system. Forty fractions were collected and consecutive fractions were combined as before for ease of immunoblotting. The fractions were immunoblotted for MRFAP1, MORF4L1 and MRGBP in addition to histone H3 and tubulin (Appendix Figure 1.1). Comparison of the two sets of immunoblots demonstrated a very close correlation in the results indicating that the SEC separation of the proteins from tissues is reproducible. Interestingly, in contrast to the U2OS human cell line there is a greater overlap between MORF4L1 and MRGBP in the smaller molecular weight fractions, with only very weak MRFAP1 bands seen in these fractions. It is known that MRGBP is not widely expressed in tissues and that its strongest staining pattern is seen in the testis. Therefore, the altered MRGBP profile in the SEC fractions may reflect this. This experiment demonstrates that the native protein complex SEC approach could be applied to study changes in protein complexes, in an organ specific set-up.



Appendix Figure 1.1. Native complex separation by SEC in murine testes.

Murine testes from mice aged 160 days (top panel) and 48 days (bottom panel) were prepared for native SEC separation. Consecutive elution fractions were combined for immunoblotting. 10μg of protein was loaded in the lane labelled "total", representing the material injected onto the SEC column. Up to 10μg of protein was loaded in each of the remaining lanes.

Appendix II: Publications arising from this thesis

Research

⌘ *Author's Choice*

Characterization of Native Protein Complexes and Protein Isoform Variation Using Size-fractionation-based Quantitative Proteomics*[S]

Kathryn J. Kirkwood‡§, Yasmeen Ahmad‡§, Mark Larance‡§, and Angus I. Lamond‡¶

Proteins form a diverse array of complexes that mediate cellular function and regulation. A largely unexplored feature of such protein complexes is the selective participation of specific protein isoforms and/or post-translationally modified forms. In this study, we combined native size-exclusion chromatography (SEC) with high-throughput proteomic analysis to characterize soluble protein complexes isolated from human osteosarcoma (U2OS) cells. Using this approach, we have identified over 71,500 peptides and 1,600 phosphosites, corresponding to over 8,000 proteins, distributed across 40 SEC fractions. This represents >50% of the predicted U2OS cell proteome, identified with a mean peptide sequence coverage of 27% per protein. Three biological replicates were performed, allowing statistical evaluation of the data and demonstrating a high degree of reproducibility in the SEC fractionation procedure. Specific proteins were detected interacting with multiple independent complexes, as typified by the separation of distinct complexes for the MRFAP1-MORF4L1-MRGBP interaction network. The data also revealed protein isoforms and post-translational modifications that selectively associated with distinct subsets of protein complexes. Surprisingly, there was clear enrichment for specific Gene Ontology terms associated with differential size classes of protein complexes. This study demonstrates that combined SEC/MS analysis can be used for the system-wide annotation of protein complexes and to predict potential isoform-specific interactions. All of these SEC data on the native separation of protein complexes have been integrated within the Encyclopedia of Proteome Dynamics, an online, multidimensional data-sharing resource available to the community. ***Molecular & Cellular Proteomics 12: 10.1074/mcp.M113.032367, 3851–3873, 2013.***

The majority of proteins function as part of multiprotein complexes and not as isolated polypeptides. These protein complexes range from simple homodimers to large structures composed of many different polypeptides. Protein complexes vary in their size and shape from small globular dimers, such as 14-3-3 proteins, to large elongated filaments of variable length, such as microtubules. The wide variety of possible protein–protein interactions within multiprotein complexes contributes to the diversity of functions that are involved in cellular processes and regulatory mechanisms.

Another important source of functional diversity and regulation is the large number of protein isoforms that may be generated from each gene. Functionally and structurally distinct isoforms can arise via multiple mechanisms, including alternative splicing, post-translational modification (PTM),[1] and proteolytic cleavage. Distinct isoforms can exhibit radically different properties. For example, including or excluding individual exons can either create or remove protein–protein interaction interfaces for binding specific interaction partners. Similarly, phosphorylation, and other PTMs, can either create or remove binding sites for interacting proteins, substrates, or ligands. PTMs can also promote structural changes in proteins and affect catalytic activity.

The association of protein isoforms and post-translationally modified factors in multiprotein complexes can influence their subcellular location, activity, and substrate specificity. This can be dynamically regulated to modulate protein complex composition, and hence localization and function, to allow cells to respond to spatial and temporal stimuli. It is therefore important to characterize protein complexes at the level of the protein isoforms and post-translationally modified forms they contain in order to fully decipher the network of signaling and regulatory pathways within cells.

Although many types of protein complexes have been studied in detail, in-depth analysis of the composition, dynamics, and isoform association of protein complexes formed in either

From the ‡Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow St., Dundee, DD1 5EH, United Kingdom

⌘ *Author's Choice*—Final version full access.

Received July 4, 2013, and in revised form, August 24, 2013

Published, MCP Papers in Press, September 16, 2013, DOI 10.1074/mcp.M113.032367

Author contributions: K.K. and M.L. performed SEC experiments. K.K. performed the immunoblotting. Y.A. performed the statistical analysis and created the Encyclopedia of Proteome Dynamics. K.K., Y.A., M.L., and A.I.L. wrote the paper. A.I.L. mentored the project.

[1] The abbreviations used are: EPD, Encyclopedia of Proteome Dynamics; GO, gene ontology; HP1BP3, heterochromatin protein 1–binding protein 3; IP, immunoprecipitation; LDS, lithium dodecyl sulfate; MORF4L1, mortality factor 4-like protein 1; MRFAP1, MORF4 family-associated protein 1; MRGBP, MRG/MORF4L-binding protein; PTM, post-translational modification; SEC, size-exclusion chromatography; TCEP, triscarboxyethylphosphine.

human cells or model organisms is still not well documented at a system-wide level. The CORUM database, compiled using a variety of information from the literature describing protein interactions and assemblies, currently provides the largest public dataset of protein complexes (1). CORUM contains information relating to ~1,970 protein complexes identified in human cells. However, these complexes are formed from proteins encoded by only ~16% of the known human protein-coding genes, indicating that many forms of protein complexes still remain to be identified and characterized (1). Furthermore, the CORUM database does not describe how the protein compositions of the complexes may vary, either dynamically or in different subcellular locations, or how this relates to protein isoforms and PTMs. This illustrates that there is still a major deficit in our knowledge of the structure and functions of cellular protein complexes and how they contribute to biological regulatory mechanisms.

The technique that is now most widely used to identify the components of protein complexes is affinity purification of an individual "bait" protein and subsequent analysis of the co-isolated proteins, usually via mass spectrometry (2). Affinity purification can use antibodies specific for an endogenous target protein (3, 4), if available, or, alternatively, can utilize a genetically constructed, epitope-tagged bait protein. The latter procedure is now widely used and is advantageous in that many different complexes can be compared using an identical antibody, or other affinity-purification method, targeted to the tag on the bait (for examples, see Refs. 5–7; for reviews, see Refs. 8 and 9). In contrast, it is harder to directly compare the results from immunoprecipitation of different endogenous protein complexes because each specific antibody that is used has different affinities and properties. Nonetheless, although epitope-tag "pull-down" techniques are now commonly used, they also have limitations. Not least, the addition of epitope tags to the bait can affect protein function and interactions (10, 11).

To help determine whether co-purifying proteins detected using pull-down strategies represent specific partner proteins in *bona fide* complexes or are nonspecific contaminants, we and others have developed quantitative approaches, for example, based on variations of the stable isotope labeling of amino acids in cell culture (12–16). Additional data analysis procedures, including the use of a "super experiment" database that predicts the likelihood of nonspecific protein interactions based on the frequency with which any given protein is co-purified across many separate experiments, can also help to define the composition of protein complexes (17). Nonetheless, affinity purification strategies have a limited ability to distinguish multiple related complexes that may differ with respect to isoforms and PTMs. They are also costly and difficult to implement for large-scale studies to survey cellular complexes, and thus not well suited to study variations in complexes under different cellular growth conditions and responses.

For system-wide studies of the composition and dynamics of protein complexes, alternative methods, in addition to immune-affinity purification, are required for convenient separation, characterization, and comparison of cellular protein complexes. To address this, a number of studies have utilized various forms of either column chromatography or native gel electrophoresis in combination with mass-spectrometry-based proteomics. For example, protein complexes have been separated using techniques including blue native polyacrylamide gel electrophoresis (18, 19), ion-exchange chromatography (20), and size-exclusion chromatography (21, 22) prior to MS analysis of proteins in the fractionated complexes. Size-exclusion chromatography (SEC) is a well-established technique used to separate proteins and protein complexes in solution on the basis of their shape/size (rotational cross-section) (23). SEC has been extensively used as an intermediate step in conventional multistep biochemical protein purification strategies. In contrast, SEC has been less commonly combined with mass-spectrometry-based proteomics for the high-throughput characterization of protein complexes. However, this has been demonstrated in previous studies that analyzed native protein complexes in plant chloroplasts (22) or large cytosolic complexes in mammalian cells (21).

In this study, we combined native SEC with high-throughput mass-spectrometry-based proteomic analysis to characterize soluble protein complexes isolated from human osteosarcoma cells. Herein we demonstrate the utility and reproducibility of this approach for the system-wide characterization of endogenous, untagged protein complexes and show how it can be used to identify specific protein isoforms and PTMs associated with distinct protein complexes. The resulting data are available to the community in a convenient format in the Encyclopedia of Proteome Dynamics (EPD) (www.peptracker.com/encyclopediaInformation/), a user-friendly, searchable online database.

EXPERIMENTAL PROCEDURES

Materials—U2OS cells were purchased from the American Type Culture Collection (Rockville, MD). Dulbecco's modified Eagle's medium (DMEM), fetal calf serum, antibiotics, NuPage gels, LDS sample buffer, MES SDS-PAGE running buffer, nitrocellulose iBlot stacks, SYPRO Ruby, Alexa Fluor 680-conjugated secondary antibodies, Dulbecco's phosphate-buffered saline (PBS), EZQ protein quantitation reagent, and a CBQCA assay kit were obtained from Invitrogen (Carlsbad, CA). IrDye 800-conjugated secondary antibodies were obtained from Rockland Immunochemicals (Gilbertsville, PA). Histone H3, NEDD8, and CUL3 primary antibodies and HRP conjugated secondary antibodies were from Cell Signaling Technology (Danvers, MA). Bicinchoninic acid (BCA) assay reagents, Coomassie Plus (Bradford) reagent, a Subcellular Protein Fractionation Kit, detergent removal plates, Acclaim Pepmap C18 columns and trapping cartridges, and triscarboxyethylphosphine (TCEP) (bond-breaker neutral pH solution) were from Thermo Scientific (Waltham, MA). Trypsin Gold was from Promega (Madison, WI). Sep-Pak tC18 96-well μ-elution plates were from Waters (Milford, MA). GAPDH primary antibody, complete protease inhibitor mixture tablets, and PhosStop phosphatase inhibitor tablets were from Roche (Basel, Switzerland). Odyssey nitrocellulose membrane was from Li-Cor Biosciences (Lincoln, NE). MRFAP1 pri-

mary antibody was from Proteintech Group (Chicago, IL). MORF4L1 primary antibody was from Santa Cruz Biotechnology Inc. (Dallas, TX). MLN4924 compound was a kind gift from Professor Sir Philip Cohen. PAF primary antibody was from Abcam (Cambridge, UK). MG132 was from Calbiochem-Novabiochem (Nottingham, UK). Ultrafree-MC centrifugal filter units were from Millipore (Billerica, MA). All other materials were obtained from Sigma (St. Louis, MO).

Cell Culture—Briefly, U2OS cells were grown in DMEM supplemented with 10% FCS, 100 U/l penicillin, and 100 μg/l streptomycin at 37 °C in 10% CO_2 and passaged at ~80% confluence. U2OS cells expressing LAP1-tagged MRFAP1 were grown in the same medium but with the addition of 150 μg/ml hygromycin B and 15 μg/ml blasticidine HCl.

GFP-IP from LAP1-MRFAP1 U2OS Cell Line—The cells for each condition were harvested separately via trypsinization, washed in PBS, and lysed in IP buffer (1% Nonidet P-40, 50 mM Tris-HCl, pH 7.4, 10% glycerol, 150 mM NaCl, complete protease inhibitor mixture (Roche Applied Science), PhosStop, 50 mM N-ethylmaleimide). The lysates were sonicated for 10 s at 10% power (three times in total) and then centrifuged for 10 min at 17,000*g* at 4 °C. Equal protein amounts of each sample were then combined with GFP-trap agarose beads from ChromoTek (Martinsried, Germany) that had been washed once in IP buffer (40 μl of 50% GFP-trap bead slurry per IP) and incubated for 2 h at 4 °C with rotation. The beads were then washed three times with IP buffer via centrifugation at 2,000*g* for 2 min at 4 °C. The beads were resuspended in 200 μl of PBS, transferred to a spin column (Thermo Scientific), and centrifuged dry at 500*g* for 1 min. LDS sample buffer that had been preheated to 65 °C was then added and incubated at 65 °C for 5 min. The eluate was collected via centrifugation at 500*g* for 1 min at room temperature.

Lysis of Cells for SEC—Ten 15-cm dishes (80% confluent) of U2OS cells were scraped, on ice, in 500 μl of ice-cold PBS containing Complete Protease Inhibitors EDTA-free (Roche) and PhosStop (Roche). Total cell lysates were sonicated for 10 s, three times in total, at 10% power at 4 °C and then centrifuged at 17,000*g* for 10 min at 4 °C. Samples were filtered through 0.45-μm Ultrafree-MC centrifugal filter units (Millipore). Bradford assays were performed on the filtrates for protein quantitation.

SEC, Trypsin Digestion, and Peptide Clean-up—Using a Dionex Ultimate 3000 HPLC system (Thermo Scientific), lysates in PBS and inhibitors were injected (200 μl per injection) onto a Superose 6 10/300GL column (GE Life Sciences) equilibrated with PBS (pH 7.2). The flow rate was 0.2 ml min^{-1}, and 40 200-μl fractions were collected using a low protein binding 96-deep-well plate Eppendorf (Hamburg, Germany). Tris-HCl (1 M pH 8.0) was added to each fraction to a final concentration of 0.1 M to adjust the pH to 8.0, and trypsin diluted in 0.1 M Tris-HCl was added at ratios of 1:50 to 1:100. The fractions were then incubated for 18 h at 37 °C. For peptide desalting, trifluoroacetic acid (TFA) was added to a 1% (v/v) final concentration and peptides were purified using a Sep-Pak tC18 96-well μ-elution plate. Peptides were eluted in 200 μl of 50% (v/v) acetonitrile, 0.1% TFA, and a SpeedVac was used to dry them prior to resuspension in 5% (v/v) formic acid. Peptide concentrations were determined using the CBQCA assay after 25-fold dilution of peptide samples in 0.1 M borate buffer, pH 9.3.

Denaturing Size-exclusion Chromatography—For the denaturing SEC, cells were lysed in PBS containing inhibitors as above and supplemented with 4% SDS, 25 mM TCEP, and 50 mM N-ethylmaleimide. The column was equilibrated with 0.2% SDS, 100 mM NaCl, and 10 mM $NaPO_4$. The lysate was otherwise injected onto the Superose 6 column as described above. Consecutive elution fractions were combined, heated to 65 °C for 10 min, subjected to chloroform methanol precipitation (24), and resuspended in 1X LDS with 25 mM TCEP for immunoblotting.

SDS-PAGE and Immunoblotting—For immunoblotting of non-denatured samples, Bradford protein quantitation assays were performed on the fractions. 20% SDS was added to each fraction to a 2% final concentration, and fractions were heated to 65 °C for 10 min. 100 μl of consecutive fractions were combined, and chloroform methanol precipitation was performed (24). Protein was then resuspended in equal volumes of 1X LDS, 25 mM TCEP so that the maximum concentration in the most concentrated fraction was 1 mg/ml and heated to 65°C for 10 min. Combined fractions were analyzed via EZQ quantitation assay (Invitrogen). 10 μl of each fraction was loaded per lane for SDS-PAGE. BCA protein quantitation (Thermo Scientific) was performed on denatured samples. Equal volumes (14 μl) of consecutive samples were combined and made up to a maximum of 0.1 mg/ml in 1X LDS/TCEP. 20 μl of sample was loaded per lane for SDS-PAGE. SDS-PAGE was performed using 4–12% (w/v) Bis-Tris NuPage gels using MES running buffer according to the manufacturer's instructions, but with the addition of 25 mM TCEP in the LDS sample buffer. SYPRO Ruby staining was performed as per the manufacturer's instructions (Invitrogen). For Western blotting, separated proteins were electrophoretically transferred to either an iBlot nitrocellulose membrane or an Odyssey nitrocellulose membrane, blocked with 3% nonfat milk in 0.1% Tween-20 in TBS, and incubated with primary antibody in 5% BSA in TBS Tween 20 overnight at 4 °C. After incubation, membranes were washed three times in TBS Tween 20 and incubated with either HRP-labeled or Alexa fluor 680/IrDye 800–labeled secondary antibodies in 3% nonfat milk in TBS Tween 20. Proteins were visualized using Immobillon chemiluminescent substrate (Millipore) and imaged with either a cooled charge-coupled device camera GE Life Sciences (Piscataway, NJ) for HRP-labeled secondary antibodies or a Licor Odyssey CLx imager for Alexa fluor 680/IrDye 800–labeled secondary antibodies.

LC-MS/MS and Analysis of Spectra—Using a Thermo Scientific Ultimate 3000 nanoHPLC system, 1 μg of peptides in 5% (v/v) formic acid (~10 μl) was injected onto an Acclaim PepMap C18 nano-trap column Thermo Fisher Scientific (Waltham, MA). After being washed with 2% (v/v) acetonitrile, 0.1% (v/v) formic acid, peptides were resolved on a 150 mm × 75 μm Acclaim PepMap C18 reverse-phase analytical column over a gradient from 2% acetonitrile to 80% acetonitrile over 100 min with a flow rate of 300 nl min^{-1}. The peptides were ionized by nano-electrospray ionization at 1.2 kV using a fused silica emitter with an internal diameter of 5 μm from New Objective (Woburn, MA). Tandem mass spectrometry analysis was carried out on an LTQ-Velos Orbitrap mass spectrometer (Thermo Scientific) using collision-induced dissociation fragmentation of precursor peptides and fragment ion measurement in the linear ion trap. The data-dependent acquisition method used was the FT10 protocol, described elsewhere (25). The RAW data produced by the mass spectrometer were analyzed using the quantitative proteomics software MaxQuant (26) (version 1.3.0.5). This version of MaxQuant includes an integrated search engine, Andromeda (27). The database supplied to the search engine for peptide identifications was the human UniProt database (June 7, 2011) containing 109,824 entries. The mass tolerance was set at 6 ppm for precursor ions, and the MS/MS mass tolerance was set at 0.5 Da. The enzyme was set as trypsin with up to two missed cleavages. Deamidation of Asn and Gln, oxidation of Met, pyro-Glu (with N-term Gln), and phosphorylation (STY) were set as variable modifications. N-ethylmaleimide on Cys was searched as a fixed modification. Identification was set to a false discovery rate of 1%. To achieve reliable identifications, all peptides identified were accepted based on the criteria that the number of hits in the "forward" database was at least 100-fold greater than the number of "reverse" database hits, thus resulting in a false discovery rate of less than 1%. The output from MaxQuant provided peptide level data as well as protein level data, grouped by protein isoforms.

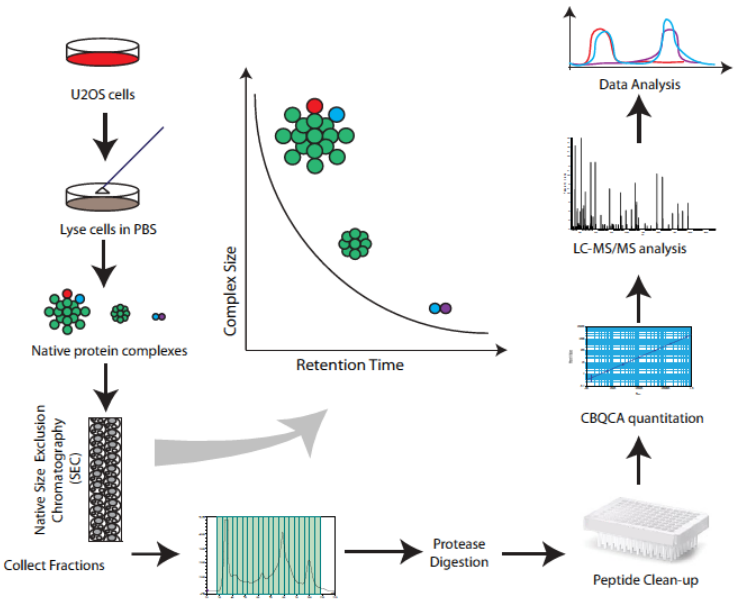


FIG. 1. **Native protein complex analysis using a comprehensive proteomics workflow.** Workflow for SEC-based protein complex separation.

Data Analysis—To create an elution profile for an individual protein isoform in each of the three replicates, the number of peptides observed in each fraction was summed to generate a count of peptides per fraction (spectral counts). The resulting spectral count profile minima and maxima were normalized within the limits of 0 and 1, respectively. The average of the three elution profiles was based on the normalized spectral peptide counts for each fraction. A minimum of two peptides were required for inclusion in our analysis. Further downstream data interpretation for this study was performed primarily using the R language (version 2.15.1). From the three biological replicates, it was required that a protein be identified in at least two replicates with a minimum of two peptides in each. Proteins labeled as either contaminants or reverse hits were removed from the analysis. These stringent criteria left a subset of 8,165 proteins with high data coverage across replicates (73.2%). For each protein, 120 quantitative measurements were recorded across the three replicates and 40 SEC fractions. The first six fractions were deemed to contain complexes outside the resolution range of the column and thus were excluded from further data analysis.

To analyze reproducibility between replicates, the peptide count profile across the remaining 34 fractions for each replicate was correlated between replicates. A quality filter was applied to ensure that a minimum of two out of the three pairwise correlations had a positive correlation coefficient. The biological replicates were collapsed by averaging to obtain a resultant mean peptide count profile describing the behavior of each protein across the 34 fractions. The standard deviation for the three biological replicates per fraction was also calculated to provide standard error bars on the protein profile graphs. The protein count profiles across the 34 fractions were normalized using the maximum peptide count for that protein and then hierarchically clustered based on Euclidean distance measurement and a "complete" agglomeration method. The output of the clustering has been presented in a heatmap using the RColorBrewer library.

The gene ontology analysis was carried out using the DAVID Functional Annotation Tool for Biological Processes (28). The full *Homo sapiens* proteome, supplied by DAVID, was selected as a background list. When cross-analyzing with the CORUM dataset, a set proportion of proteins in a complex were required to co-fractionate to provide sufficient evidence that a complex had been identified. To calculate this proportion, we used the equation max(0.3, $2/n$) × 100, meaning 30% of the components or 2 divided by the number of proteins in the complex, whichever was greater, multiplied by 100. This approach ensured that for small complexes (*i.e.* containing six or fewer components), more than simply one or two of the component proteins needed to co-fractionate before we labeled the complex as identified.

RESULTS

Workflow for the Identification of Native Protein Complexes—To prepare lysates for the characterization of native protein complexes from human U2OS osteosarcoma cells, cellular lysis was performed by resuspending cells in PBS buffer containing protease and phosphatase inhibitors and then subjecting samples to sonication. PBS buffer provides a neutral pH and a physiological salt concentration that mildly solubilizes most protein complexes. No detergents were used to minimize the disruption of protein–protein interactions (Fig. 1). The resulting protein complexes were fractionated by size/shape using non-denaturing SEC. The eluate was collected in 40 sequential fractions of equal volume (Fig. 1). For proteomic analysis, the fractions were digested with trypsin, and then the peptides were desalted and analyzed via LC-MS/MS. Peptides were identified using MaxQuant (26) and quantified by the normalized spectral count in each fraction with standard parameters (see "Experimental Procedures"). Three biological replicates were performed and analyzed identically. In total, over 71,500 peptides, corresponding to over 8,100 proteins, were identified after the data had been filtered such that each protein was identified by at least two peptides in at least two out of the three biological replicates (supplemental Tables S1 and S2).

Protein standards were analyzed using our SEC method to identify the approximate molecular weight range across each fraction. These standards are ideal globular proteins and so provide a general size indication. However, SEC also separates proteins and their complexes by shape (rotational cross-section). Therefore, an elongated protein complex can appear larger than its true molecular weight. This aside, our analysis showed that overall this method separated protein complexes with estimated molecular weights ranging from ~15 kDa to >1 MDa (Fig. 2*A*).

The elution profile for each protein standard was used to calculate a predicted molecular weight range for each SEC fraction, from which we generated a linear regression model (supplemental Fig. S1). This model was then used to calculate the approximate observed molecular weight of all complexes/proteins identified in our dataset based on their peak elution fraction. The predicted molecular weights of all proteins identified through MS analysis were extracted from the UniProt database (29) and compared with the observed molecular weight, as estimated from the linear regression model, for the fractionation behavior of each protein (supplemental Fig. S2). The resulting scatter plot shows that most proteins were identified fractionating from SEC at an apparent molecular weight greater than the predicted molecular weights of individual proteins. This strongly supports our hypothesis that with cell extracts prepared using physiological buffer conditions and with SEC carried out using native conditions, the majority of proteins in the human proteome are detected in association with one or more complexes.

To evaluate whether our SEC method predominantly separated intact protein complexes, an analysis was carried out in which pairs of consecutive elution fractions were combined, resulting in 20 combined fractions in total. The pairwise combined fractions were further separated by molecular weight, using SDS-PAGE, and the resulting gel was stained for total protein (Fig. 2*B*). This showed that the earliest to middle fractions eluting from the SEC column (corresponding to larger complexes) contained proteins with very heterogeneous sizes, whereas fractions eluting progressively later (corresponding to smaller complexes) contained a greater proportion of proteins that migrated in SDS-PAGE close to their predicted molecular weight (Fig. 2*B*). This suggests that small complexes are less heterogeneous, as would be expected, and that monomeric proteins that do not form complexes are mostly found in these fractions. We also observed that the first three fractions (labeled "void") included proteins that likely form complexes larger than the separation range of the SEC column (Fig. 2*B*).

As a control, we also analyzed lysates from U2OS cells prepared using a highly denaturing (4% SDS-containing) PBS buffer, which will disrupt essentially all non-covalent protein complexes. In this case there was a major shift in the elution profiles of most proteins from the SEC column relative to the previous analysis of lysates under native conditions. Even though denatured proteins in general will appear larger than their native forms because of the loss of their globular structure, this appears to have only a relatively small effect on elution behavior. When the highly denaturing lysate buffer was used, most proteins migrated at, or close to, their predicted molecular weight when injected into the same SEC column (supplemental Fig. S3). From these data we conclude that the native SEC analysis separates *bona fide* multiprotein complexes across a broad range of fractions.

Reproducibility of the SEC Elution Profiles—To determine the reproducibility of protein complex separation by SEC, we performed a pairwise comparison of the proteomic analysis of all proteins in each dataset across each of the three biological replicates (*i.e.* 1 *versus* 2, 2 *versus* 3, 1 *versus* 3). In each comparison a Pearson correlation coefficient was calculated to compare each normalized protein elution profile across the 40 SEC fractions. Each pairwise comparison is shown in a density plot that reveals a high correlation (>0.8) between the replicates for the majority of proteins (Fig. 3*A*). Most proteins demonstrated one or more peaks in their profile, as illustrated in the elution profiles for a representative group of proteins (Fig. 3*B*). We hypothesized that many of these peaks will correspond to the elution of proteins that are, at least in part, associated in complexes with other proteins that co-elute within that fraction.

Many proteins showed complex elution profiles, as expected, because we were detecting within the cell lysate a combination of protein polymers, large and small complexes as well as monomeric proteins. To characterize further the separation of protein complexes by native SEC, we analyzed the co-fractionating proteins using the orthogonal technique of Western blotting. For this, the pairwise pooled SEC fractions, as described above, were separated via SDS-PAGE and immunoblotted to detect four representative proteins with different elution profiles displaying profile peaks in different fractions (Fig. 3*B* and supplemental Fig. S4). Regardless of whether the protein was a small ubiquitin-like modifier (NEDD8), a protein involved in large complexes (CUL3), a protein known to form either multimers (GAPDH) or extended polymers (tubulin), or a small DNA damage repair protein (PAF), we saw not only a close match within the MS replicate data (*i.e.* relatively small error bars), but also a close match between the MS data and the immunoblot profiles.

In summary, all of the data comparing the SEC fractionation of proteins by MS-based analysis and immunoblotting showed that the SEC approach was reproducible and demonstrated that both detection methods gave similar results that were in agreement with previously published information.

Analysis of Protein and Complex Coverage—To determine whether the resulting SEC protein dataset had any intrinsic bias for the preferential detection of either low or high copy number proteins, we compared our dataset with the CORUM database, which records known protein complexes (1). For this comparison we used previously published information

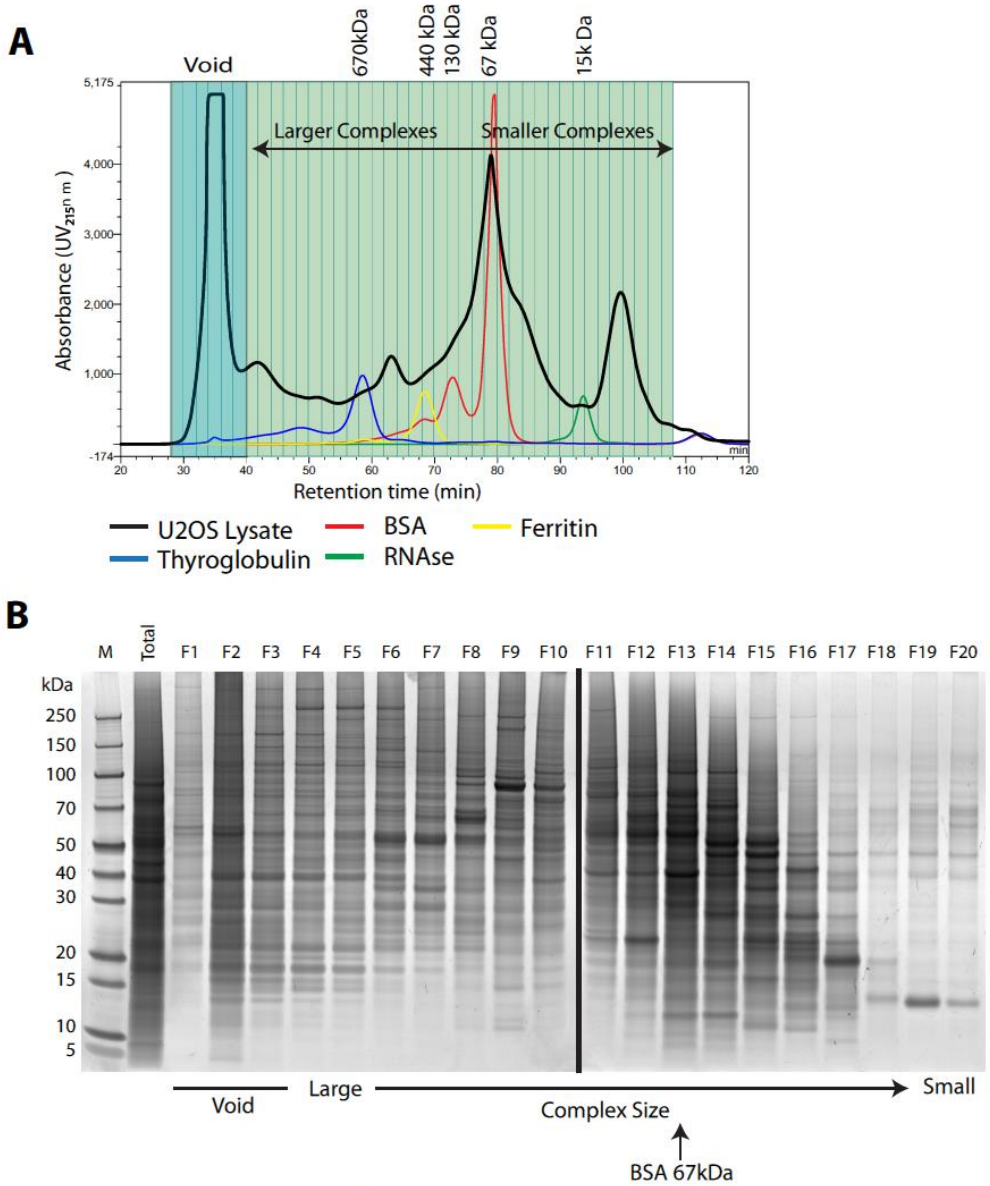


FIG. 2. **Characterization of protein complex separation.** *A*, the UV chromatogram from one of the biological replicates indicates the profile of the U2OS total cell lysate as it eluted from the Superose 6 SEC column across 40 fractions. The retention time, measured in minutes, is represented on the *x*-axis. UV absorbance is represented on the *y*-axis. Protein standards of known molecular weights (thyroglobulin, ferritin, BSA, and RNase) were injected onto the same column, and their elution peaks were used to calculate approximate molecular weights for fractions. Proteins eluting in fractions 1–6 were termed the void. Complexes that eluted in fractions 7–23 were classed as "larger," and those in fractions 24–40 were classed as "smaller." *B*, SYPRO ruby total protein stain of proteins in the total cell lysate and consecutively combined fractions. 10 μg of protein were loaded in lane 2 (total), and a maximum of 10 μg of protein were loaded per lane for elution fractions.

reporting protein abundance levels in human cells (30). This showed that there is a wide distribution of copy numbers for proteins found in complexes reported in the CORUM database (Fig. 4*A*). The same comparison of protein copy numbers applied to our U2OS cell SEC dataset demonstrated a related distribution, but with more representation of the higher copy

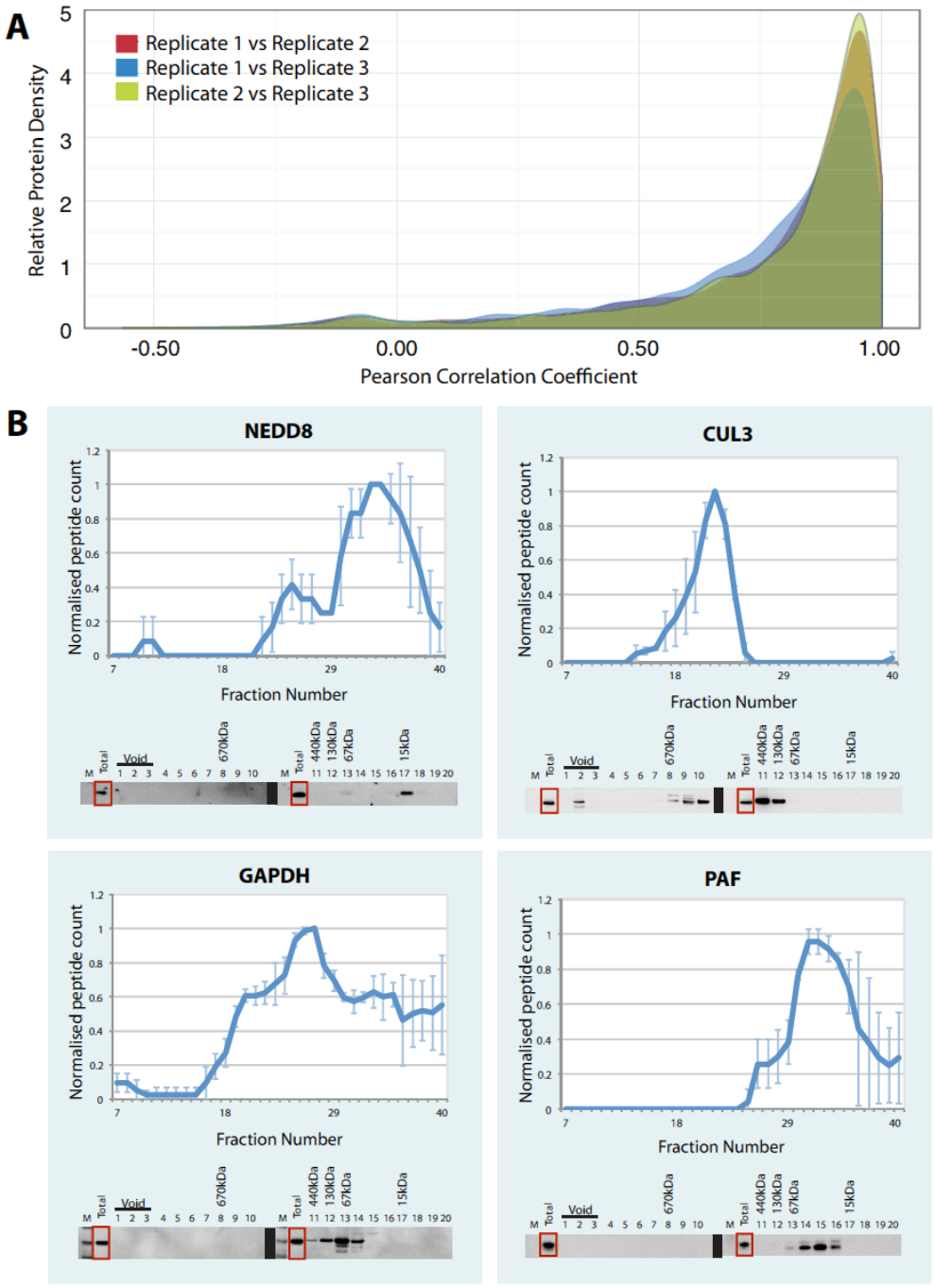


FIG. 3. **Reproducibility of the SEC elution profiles.** *A*, the Pearson correlation coefficient was calculated for the three biological replicates (replicate 1 *versus* replicate 2, replicate 1 *versus* replicate 3, and replicate 2 *versus* replicate 3). The relative protein density (*y*-axis) was plotted against the Pearson correlation coefficient for each replicate (*x*-axis). *B*, the SEC elution profiles for selected proteins are displayed as line graphs with the normalized peptide count (*y*-axis) plotted against each elution fraction (*x*-axis). Error bars indicate the standard deviation from

numbers per cell protein (Fig. 4*B*). A Wilcoxon–Mann–Whitney test used to compare both distributions gave a p value of 7.006×10^{-5}, demonstrating a statistically significant difference, likely due in part to the increased representation of high copy number proteins in our dataset. Overall, our SEC dataset contained 1,464 out of the 2,713 human proteins reported in the CORUM database (*i.e.* 54%). The lower number of proteins in part reflects the fact that there are differences in the expressed proteomes between the U2OS cell line we used for this study and the multiple different cell lines used to generate the CORUM database. In addition, our SEC analysis involved the extraction of complexes using native conditions without detergent. Therefore, potentially, specific types of complexes, including membrane-associated and/or chromatin-associated proteins, may have been either underrepresented or absent. Despite this limitation in protein extraction efficiency, this study was still able to sample more than 50% of the proteins estimated to constitute the U2OS proteome (31).

To estimate how many of the complexes described in the CORUM database were identified in this study, we cross-correlated proteins that we observed co-fractionating in SEC with the complexes described in CORUM. Analysis of the CORUM dataset indicated that 74% of the annotated protein complexes contain only four components or less. Using this figure as a guide, we deemed a protein complex detected in our dataset if we identified a significant proportion of the components within one elution fraction. To ensure that our identifications were based upon suitably stringent criteria, we applied a variable cut-off for the number of components that had to be identified to reflect the total number of proteins in different types of complex. Thus, we defined a "significant proportion" as 50% or more of the components for smaller complexes with four or fewer subunits and at least 30% of the protein components for large complexes with five or more subunits (see "Experimental Procedures"). When these strict parameters were applied, our dataset identified 1,061 of the 1,970 complexes described in the CORUM database (54%) (Fig. 4*C*). As some of the CORUM complexes will likely not be expressed in U2OS cells, this represents a greater overall fraction of the protein complexes present in U2OS cells.

Comparative Analysis of Differentially Sized Complexes—For subsequent bioinformatics analysis, protein complexes were classified by size and deemed as either "larger" (fractions 7–23) or "smaller" (fractions 24–40) complexes using the cut-off between 130 and 440 kDa (Fig. 2*A*). To examine whether a bias in protein function was evident between the larger and smaller complexes, we compared the biological process gene ontology (GO) annotations (28) for the identified proteins that peaked in larger size fractions (1,720 proteins), smaller size fractions (2,280 proteins), or both size ranges (4,076 proteins) (Fig. 5*A*). The 10 most frequently associated GO terms for proteins in each of these size categories are shown. Proteins in larger complexes were primarily associated with the terms "cell division," "chromosome organization," "vesicle mediated transport," and "ribonucleoprotein complex biogenesis" (Fig. 5*C*). In contrast, proteins in smaller complexes were mostly associated instead with the terms "oxidation reduction," "coenzyme metabolic process," "phosphate metabolic process," and "hexose metabolic process" (Fig. 5*D*). Proteins that peaked in both the smaller and larger size classes were mainly associated with yet another set of terms, mostly related to gene expression and regulatory functions, including "RNA processing," "cell cycle regulation," and "splicing" (Fig. 5*B*). This unexpectedly revealed a clear relationship in the distribution of biological functions associated with protein complexes of different sizes. Thus, each size class had a distinct set of most frequently associated GO terms, while within each class the different GO terms predominantly reflected similar functions and/or processes.

Hierarchical Clustering of Protein Elution Profiles—The separation and characterization of native protein complexes across the SEC fractions were used to facilitate the identification of putative interacting proteins, because we expected proteins interacting within complexes to display similar elution profiles. To evaluate the clustering of protein elution profiles, we therefore performed hierarchical clustering across the 34 SEC fractions. To determine the number of protein clusters required, we tested the generation of 1 to 1,000 clusters using our SEC dataset (supplemental Fig. S5). For each cluster, the correlation coefficient for the proteins within that cluster was calculated to determine the optimum cluster size, based on the similarity of protein profiles (see "Experimental Procedures"). As shown in supplemental Fig. S5, generating 200 separate clusters from the SEC dataset provided an optimum classification—that is, it showed a high correlation coefficient (above 0.9) for each cluster while minimizing the subdivision of known protein complexes into separate clusters. We therefore used 200 protein clusters as the basis for further analysis.

Four examples of protein clusters were selected, including complexes separated across the SEC fractions, and these are indicated on the hierarchical clustering heat map (Fig. 6*A*, numbers 1–4). The protein components of each of these clusters were analyzed using the STRING database to identify either previously known protein–protein interactions or other relationships within the cluster (32, 33). Examination of proteins in example cluster 1 (green annotations) showed that it contained many components of the chaperonin containing TCP1 complex, a hetero-oligomeric complex ~850 to 900

the mean across three biological replicates. Below each line graph are immunoblots for each protein of interest. A total of 10 μg of protein was loaded in the "total" lane and up to a maximum of 10 μg of protein was loaded per lane for each of the fractions. The annotated molecular weights were estimated from the elution profiles of the protein standards injected onto the SEC column.

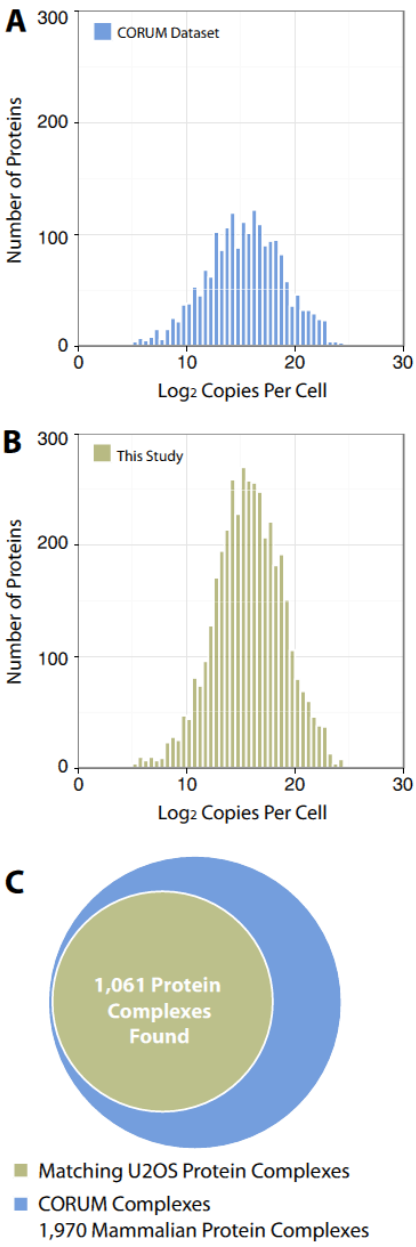


FIG. 4. **Comparison of SEC separated complexes with the CORUM database.** *A*, protein copy numbers per cell reported in a previous study (30) were applied to our SEC dataset. The results are displayed as a histogram comparing the number of proteins (*y*-axis) and the $\log_2$ copies of that protein per cell (*x*-axis). *B*, the protein copy numbers per cell were also applied to the proteins listed in the CORUM database and displayed in the same format as in part *A*. *C*, the proportion of protein complexes found in our SEC dataset that were previously described in the CORUM dataset are presented as a Venn diagram.

kDa in size (34). This correlates with the size range determined for the cognate elution fraction from the SEC column (Fig. 6*B*). In addition, elongation factor proteins that are known to interact also were present within this cluster, as were some lamins (Fig. 6*B*). Although these complexes showed their strongest elution peak around fraction 14, we were able to clearly differentiate them using a heat map and a hierarchical tree-based representation of the protein profiles (Fig. 6*C*). To further demonstrate this, we have plotted the elution profiles of each protein within example cluster 1 and highlighted the average protein profile of each labeled complex (Fig. 6*D*). This shows that small differences in the overall elution profile can be used to differentiate these separate complexes within the same cluster.

The components of the other example clusters highlighted (2–4) are illustrated in supplemental Fig. S6. Sample cluster 2 contained many components of the 80S ribosome, including proteins from both the large (60S) and small (40S) subunits. This cluster eluted from the column at a molecular weight above 670 kDa. The molecular weight of the intact ribosome is ~4 MDa (35), which is outside the effective fractionation range of the SEC column used in this study. However, we may also detect large ribosome fragments and/or assembly intermediates of ribosome subunits with an approximate molecular weight of ~1 to 2 MDa, which correlates with the observed peak in elution fraction 7 for cluster 2. Components of the eukaryotic initiation factor 3 complex were also identified within this cluster. As the STRING diagram illustrates, there are proposed interactions between the eukaryotic initiation factor 3 complex and the ribosome. This demonstrates that potentially interacting complexes can be identified within the same cluster. Components of the base unit of the 19S proteasome cap also eluted in this same cluster, demonstrating that, given the resolving capacity of the Superose 6 SEC column, multiple complexes can exist within one cluster. The 19S proteasome has a molecular weight of ~900 kDa (36), which correlates with the approximate size for this cluster.

Example cluster 3 contained several groups of known interacting proteins, including components of the DNA synthesome complex. This is a multisubunit complex involved in DNA replication (37, 38) (supplemental Fig. S6, box 3). Example cluster 4 (supplemental Fig. S6, box 4) did not contain any known large protein complex networks based on STRING analysis. Within this group, only occasional pairs of interacting proteins were identified. This result was expected because the peak elution fraction for this cluster is between 15 and 67 kDa, where we find mostly free proteins and very small complexes. The absence of major networks in this cluster when analyzed by STRING therefore supports the relevance of the interactions detected in the clusters that eluted from the SEC column in fractions that we have shown contained proteins associated within complexes.

In summary, the cluster analysis performed on the four selected examples illustrates the utility of the SEC fraction-

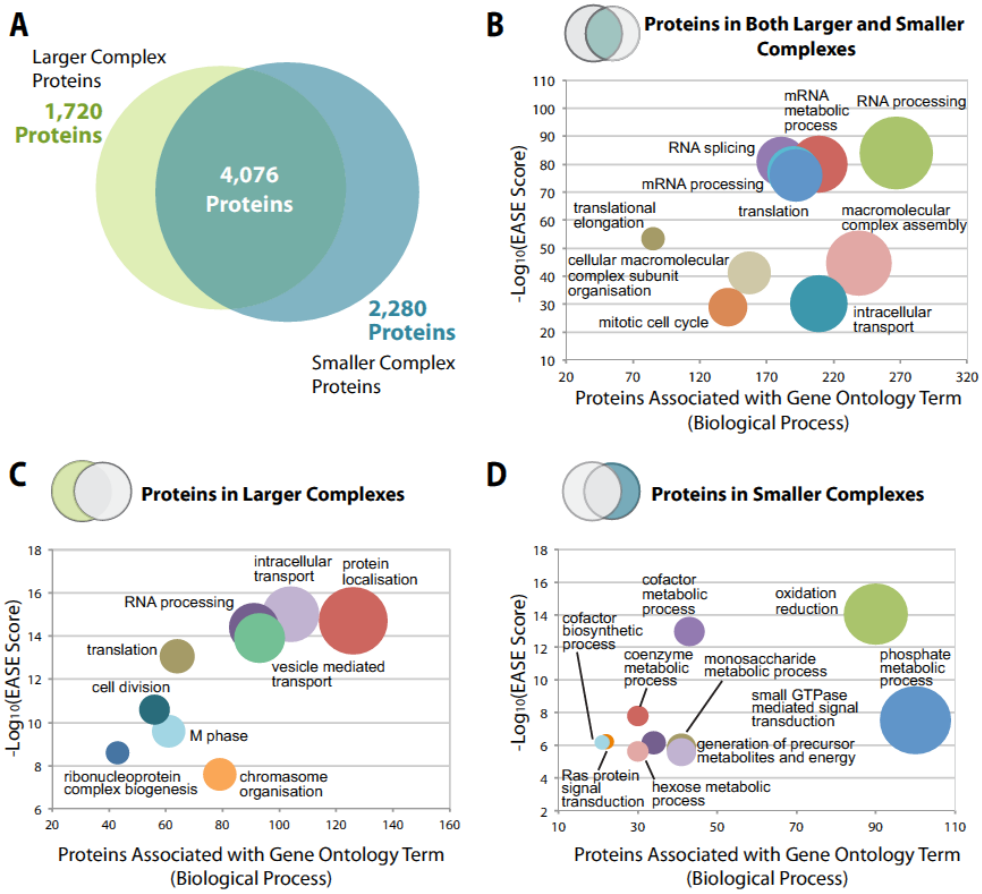


FIG. 5. **Comparative analysis of higher and lower molecular weight complexes.** *A*, the proportions of proteins identified in the SEC dataset that correspond to "larger complex proteins" (eluting in fractions 7–23) and "smaller complex proteins" (eluting in fractions 24–40) are presented as a Venn diagram. The overlap represents proteins that eluted in both larger and smaller complexes. *B–D*, bubble graphs demonstrating the top 10 gene ontology terms (biological process) (*x*-axis) plotted against the $-\log_{10}$(EASE score) for proteins eluting in both larger and smaller complexes, in larger complexes, and in smaller complexes, respectively.

ation strategy for physically separating groups of proteins based upon their interactions.

Encyclopedia of Proteome Dynamics—To provide convenient community access to the large-scale, quantitative proteomics data generated in this study, we have incorporated the entire dataset into the EPD. This is a searchable online database that collates the proteomics data from this study on protein complexes with our previous large-scale studies on subcellular protein localization and turnover in a publicly available Web-based resource (39). The integration of the multidimensional proteomics data in the EPD allows cross-correlation and analysis of protein properties. For example, this can facilitate the identification of candidate interacting proteins. As shown below, it also aids analysis of the relation of protein isoforms and PTMs to specific complexes.

To demonstrate the utility of the multidimensional data integration provided by the EPD, we analyzed the characteristics of the relatively unstudied ADP-sugar pyrophosphatase NUDT5. First, analysis of the SEC protein fractionation data in this study showed that NUDT5 is detected in at least two distinct complexes that peak in separate fractions (Fig. 7*A*). The EPD provided convenient cross-comparison with other proteins that displayed an elution profile similar to that of NUDT5, as judged by hierarchical clustering, and these data are shown in a heat map (Fig. 7*A*). Second, the EPD showed that NUDT5 is detected predominantly in the cytoplasmic fraction (Fig. 7*B*), consistent with its role in nucleotide metabolism. This result is in agreement with data based on immunofluorescence microscopy in the Human Protein Atlas, a large-scale resource reporting protein localization in U2OS

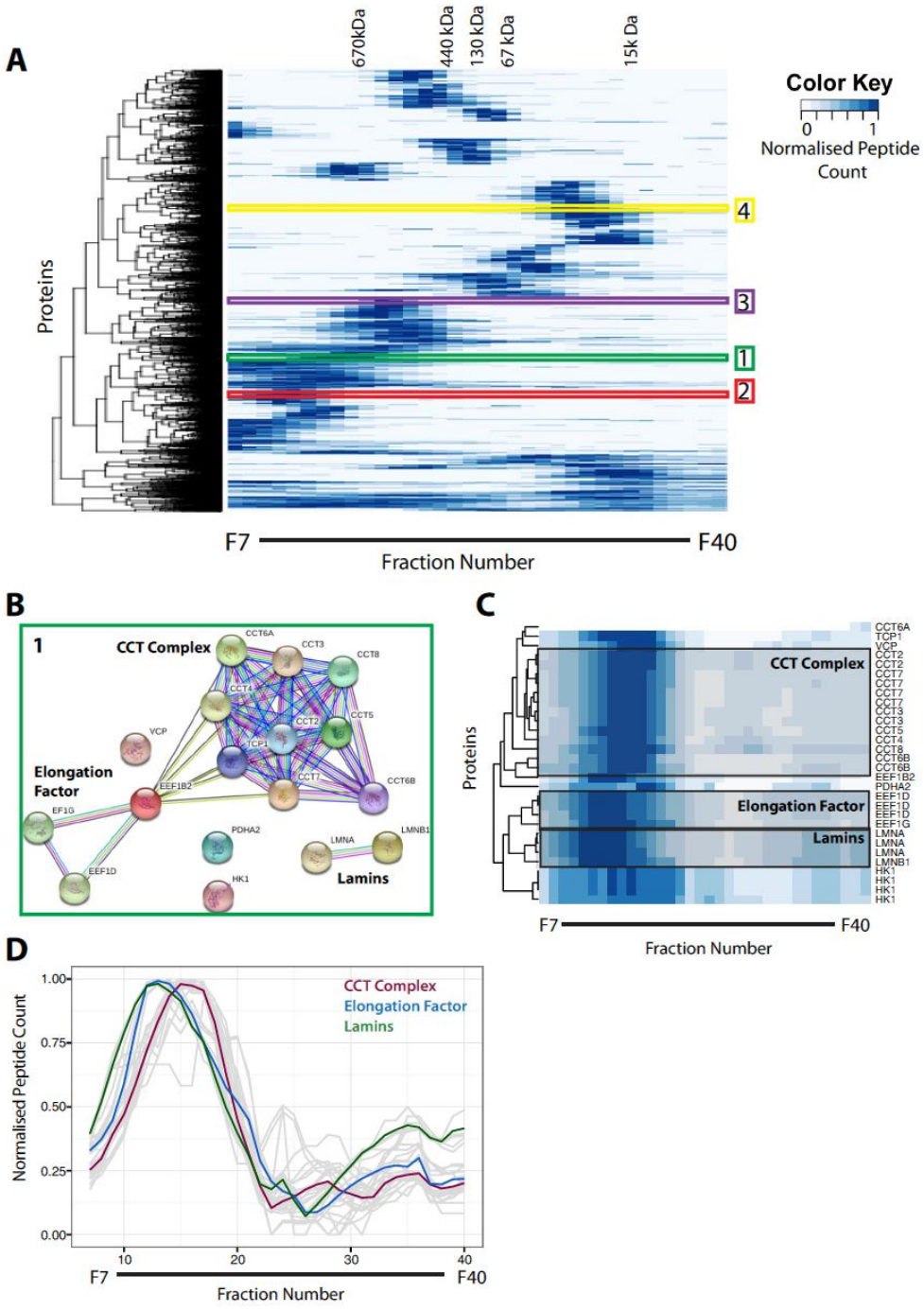


FIG. 6. **Hierarchical clustering of protein elution profiles.** *A*, all 8,165 proteins were clustered on the basis of the similarity of their normalized elution profiles across 34 SEC fractions, and the results are presented as a heat map. The dendrogram cut-offs for 200 clusters are illustrated to the left of the heat map. The peak elution fraction for each of the protein standards was used to estimate the approximate

cells and other human cell lines (40, 41). Finally, the EPD showed that NUDT5 was not among the subset of proteins with a half-life of less than six hours in U2OS cells, similar to many other metabolic enzymes, as judged by cycloheximide experiments (Fig. 7*C*).

Together, these collected measurements of protein properties that are described in the EPD for each protein will help to either predict or disprove possible interactions and potential mechanisms of regulation that can be tested in future experiments. Further network analysis and examples of phosphorylation-dependent interactions for NUDT5 are illustrated below, and the EPD facilitates similar studies on the thousands of other human proteins described in this resource.

Analysis of Proteins in Multiple Complexes—Looking more closely at the heat maps for each of the 200 protein clusters derived from the SEC elution profiles, we observed that some clusters contained proteins that eluted in two or more distinct peaks across multiple fractions. The heat map shown in Fig. 8*A* illustrates one such cluster, which includes the MORF4L1 adaptor protein (red box). MORF4L1 is known to interact with multiple proteins, including a component of the NuA4 complex called MRGBP (42), PHD factor 1, a component of the Sin3A complex (43–46), and MRFAP1 (47, 48). This explains why it showed such a broad elution profile, with multiple peaks in different fractions (Fig. 8*B*). The previous pull-down and IP data describing interactions between these proteins, however, do not reveal in detail the degree of overlap between these different MORF4L1-containing complexes.

To help dissect the MORF4L1 interactome further, we examined the elution profiles for the MORFL41 binding proteins MRGBP and MRFAP1. This demonstrated that these two proteins coincide with the MORF4L1 elution profile in distinct and nonoverlapping peaks at very different size ranges. The elution profile of MRGBP peaked in fraction 12 (Fig. 8*C*) at a molecular weight of >670 kDa, which correlates with previously published data on the size of the NuA4 complex (49). In contrast, the elution profile of MRFAP1 peaked exclusively in fractions 26–28, which corresponds to a molecular weight range between 15 and 67 kDa (Fig. 8*D*), with no peptides detected in the high molecular weight range. These data indicate the presence of at least two distinct complexes containing MORF4L1, of very different sizes and with mutually exclusive interactions with MRGBP or MRFAP1, respectively.

To complement the mass-spectrometry-based analysis, the pairs of combined SEC fractions were immunoblotted with antibodies specific for MORF4L1, MRGBP, and MRFAP1 (Fig. 9*A*). The results from the immunoblots closely correlated with the LC-MS/MS peptide data, although MRGBP was also detected in immunoblotted fractions 12 and 13. These bands likely represent monomeric MRGBP (37 kDa), as a complex containing MRFAP1, MORF4L1, and MRGBP would be predicted to elute in a higher molecular weight range. It was noted that MRFAP1 was not detected in fractions with a molecular weight above 67 kDa; therefore, it was unlikely that all three proteins formed a complex together. The SEC co-fractionation data are thus consistent with the existence of distinct MORF4L1 complexes containing either MRGBP or MRFAP1 as mutually exclusive interactive partners, as judged both by MS and immunoblotting analysis to detect the elution profiles of the respective proteins.

To test this further, we next performed IP experiments to examine whether MRGBP could be detected as a component that co-purifies in a pull-down assay with the MRFAP1–MORF4L1 complex. To do this, a stable cell line expressing wild-type, doxycycline-inducible LAP1-MRFAP1 (50) was utilized to co-immunoprecipitate proteins interacting with MRFAP1. The same immunoprecipitation assay was also performed on lysates from the same cells that had been treated with MLN4924 (1 μM for 16 h). MLN4924 is an inhibitor of NEDDylation that is known to increase the levels of MRFAP1 and MORF4L1 by reducing their degradation rates (50). In addition, the same IP assay was also performed on lysates from the same cells that had been treated with the proteasome inhibitor MG132 (10 μM for 2 h), which also increases MRFAP1 protein levels (50). Immunoblotting of these immunoprecipitates with an antibody raised against MORF4L1 confirmed that after doxycycline induction, MORF4L1 was recovered together with GFP-MRFAP1. It also showed that this co-purification was increased when cells had been treated with either MG132 or MLN4924 (Fig. 9*B*). In contrast, there were no specific MRGBP bands detected in these immunoprecipitates that could be detected using the MRGBP antibody. These data thus confirm the results from both the MS and protein blotting analyses of elution profiles in the SEC dataset.

Collectively, the data demonstrate that MRGBP–MORF4L1 and MRFAP1–MORF4L1 complexes represent distinct functional units containing mutually exclusive interaction partners. These data also validate the use of the native SEC approach to profile co-fractionation of proteins at a system-wide level and thus to predict the existence of distinct forms of protein complexes. Next, therefore, we extended our analysis of the

molecular weight of proteins/complexes eluting within that specific fraction. Four example clusters are highlighted on the heat map. The key indicates the color attributed to the normalized peptide count. *B*, STRING interaction network for proteins identified in example cluster 1. *C*, higher resolution view of the heat map for example cluster 1. Individual complexes previously identified in the literature are highlighted in gray boxes. *D*, the elution profiles for all proteins identified in example cluster 1 are illustrated as gray lines with their normalized peptide count (*y*-axis) plotted against the SEC elution fraction number (*x*-axis). The average elution profiles for co-clustering proteins known to be in highlighted complexes are overlaid in the following colors: chaperonin containing TCP1 complex in red, elongation factor complex proteins in blue, and Lamins in green.

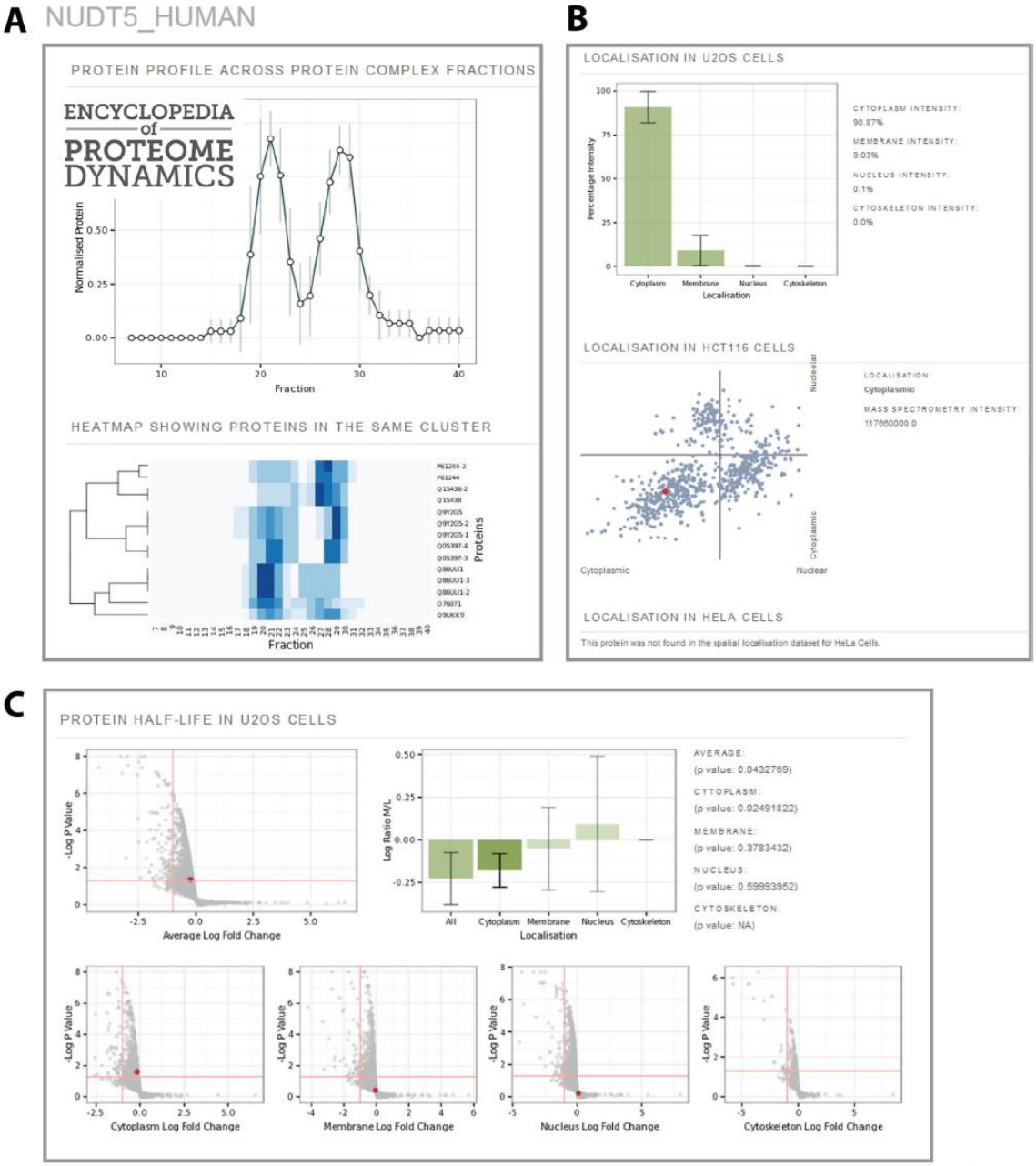


FIG. 7. **Computational analysis of large-scale experiments: the Encyclopedia of Proteome Dynamics (EPD).** *A*, screenshot from the EPD demonstrating the elution profile of the NUDT5 protein and a heat map of co-clustering proteins. The EPD also provides information regarding the subcellular localization (*B*) and turnover rates (*C*) of proteins identified in previous large-scale experiments reported by the Lamond Laboratory. The EPD database is freely accessible online.

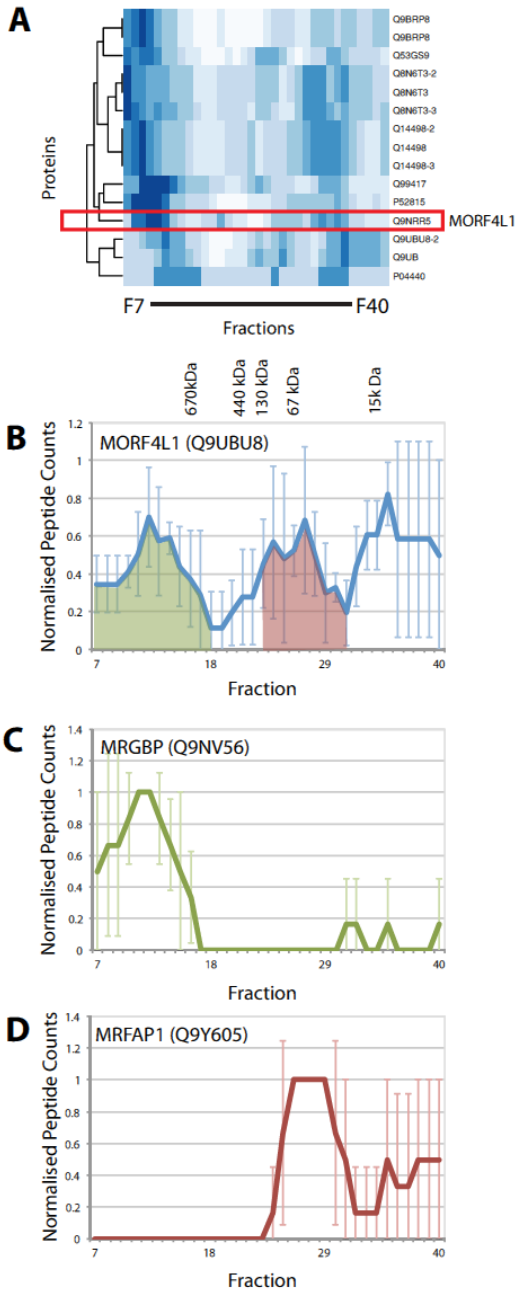


FIG. 8. **Analysis of proteins participating in multiple complexes.** *A*, heat map for one cluster containing proteins with multiple elution peaks across 34 SEC fractions. The MORF4L1 protein is highlighted in the red box. The SEC elution profiles for MORF4L1 (*B*), MRGBP (*C*), and MRFAP1 (*D*) are shown as line graphs with normalized peptide counts (*y*-axis) plotted for each SEC fraction (*x*-axis). Peaks where MORF4L1 and MRGBP profiles overlap are shaded in green, and those where MORF4L1 and MRFAP1 overlap are in red. Error bars indicate the standard deviation from the mean.

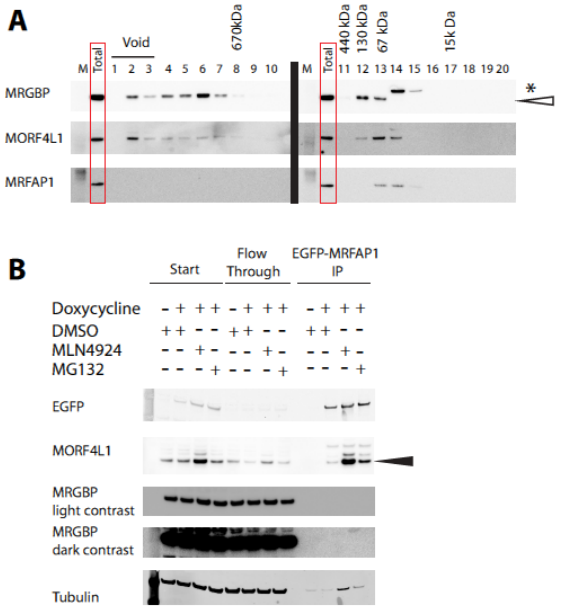


FIG. 9. **Confirmation of MORF4L1/MRGBP/MRFAP1 interactions.** *A*, consecutive elution fractions were combined for immunoblotting. 10 μg of total protein was loaded in the lane labeled "total," representing the material injected onto the SEC column. *B*, immunoblots for GFP-IP analysis of GFP-MRFAP1 cell line ($n = 3$). Cells were induced with doxycycline and treated with 1 μM MLN4924 for 16 h where indicated, 10 μM MG132 for 2 h where indicated, or dimethyl sulfoxide as a control. Solid arrow indicates the unmodified MORF4L1 protein band.

SEC dataset to profile the association of specific protein isoforms and phosphorylated proteins with distinct forms of protein complexes.

Analysis of Protein Isoforms—Mass-spectrometry-based analysis enables the unbiased detection of many protein isoforms, such as those generated by pre-mRNA alternative splice site selection, proteolytic cleavage, or protein phosphorylation. We used the known isoform annotations for the proteins identified in this study to determine whether each isoform of a single protein displayed distinct SEC elution profiles. All of the protein isoforms identified in our dataset are included in the EPD and are available for analysis. This identified examples of specific protein isoforms that may participate in different protein complexes.

The first example is the alternative splicing of heterochromatin protein 1–binding protein 3 (HP1BP3). This protein is a component of heterochromatin and has been proposed to have a role in modulating chromatin structure and function (51). HP1BP3 is known to have four isoforms formed by

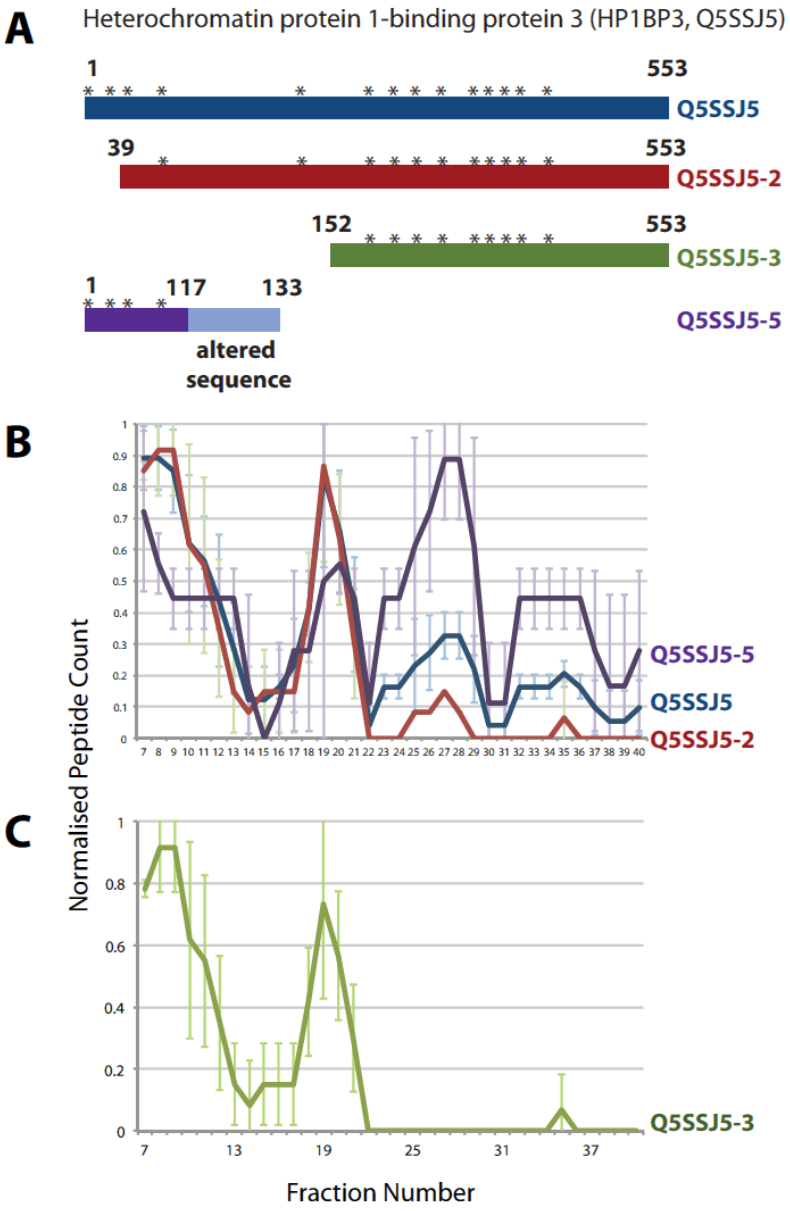


FIG. 10. **Analysis of protein isoforms formed by alternative splicing.** *A*, schematic of the four isoforms of the HP1BP3 protein. Asterisks indicate the location of identified peptides. SEC elution profiles for the canonical sequence and isoforms 2 and 5 (*B*) and the SEC elution profile for isoform 3 (*C*) are demonstrated as line graphs with the normalized peptide count (*y*-axis) plotted for each SEC fraction number (*x*-axis). Error bars denote the standard deviation from the mean.

alternative splicing, all of which were identified by up to 14 peptides spanning 26% (isoform 2) to 37% (isoform 5) of the expressed sequences of all four isoforms in our dataset (Fig. 10*A*; peptide locations are marked with asterisks). Closer analysis showed near-identical elution profiles for isoforms 1, 2, and 4 (Fig. 10*B*); isoform 3 alone was missing from a peak between fractions 22 and 30 (Fig. 10*C*). As none of the peptides were isoform specific, this discrepancy cannot be attributed to either low expression or poor sequence coverage.

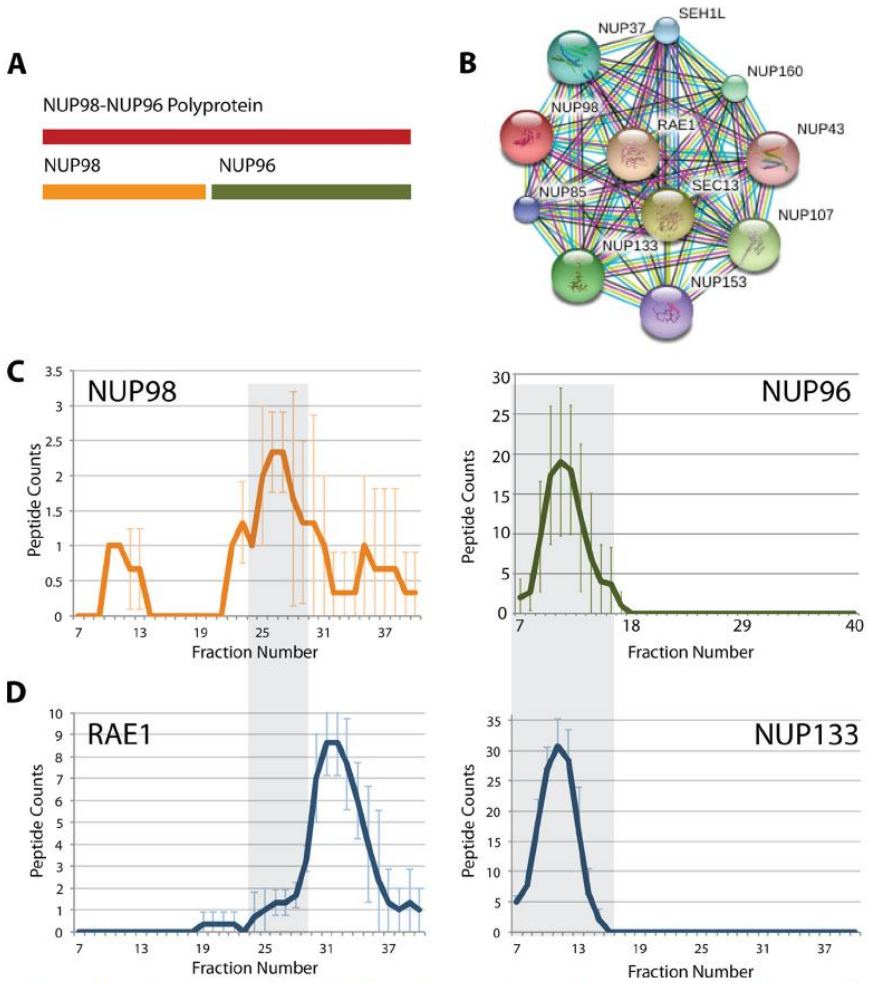


FIG. 11. **Analysis of protein isoforms formed by proteolytic cleavage.** *A*, schematic showing that the NUP98-NUP96 polyprotein is cleaved to form the separate proteins NUP98 and NUP96. *B*, NUP98 interacting proteins as proposed by the STRING database. *C*, average SEC elution profiles for peptides associated with specific regions corresponding to the NUP98 protein (orange profile) and the NUP96 protein (green profile). The line graphs represent the peptide count (*y*-axis) detected for each SEC fraction (*x*-axis). *D*, SEC elution profiles for RAE1 and NUP133 (*x*- and *y*-axes labeled as for part *C*). The gray boxes highlight fractions where we propose RAE1 and NUP98 interact and where NUP133 and NUP96 interact. Error bars indicate standard deviation from the mean.

Although little is known from the literature about the potential protein interaction partners of HP1BP3, based on our data, we predict that isoform 3 is absent from one or more protein complexes that may form with the other HP1BP3 isoforms in the molecular weight range from ~30 kDa to ~170 kDa. When the protein structures of the isoforms are compared, it is apparent that a region between positions 39 and 116 is missing in isoform 3 but is present in the other isoforms. We hypothesize that this region may be required for an interaction involved in forming the complex(es) that show an elution peak between fractions 22 and 30 (Fig. 10*A*). This will be an interesting point to pursue in the future as more information about this protein, and its isoforms, is uncovered.

In addition to protein isoforms generated by alternative splicing, we have also identified protein isoforms generated by post-translational proteolytic cleavage, such as in the NUP98 polyprotein. The NUP98 mRNA is translated into the NUP98-96 precursor protein, which is then cleaved to form the N-terminal NUP98 protein and the C-terminal NUP96 component as separate proteins (Fig. 11*A*) (52). The NUP98 pre-mRNA can also undergo alternative splicing to form only the NUP98 protein. The NUP98–96 precursor has six known

isoforms, and isoforms 3 and 4 correspond to the NUP98 protein without NUP96.

In the second example analyzed, all six known isoforms of the NUP98 protein were identified in our dataset. Interestingly, isoforms 1, 2, 5, and 6 (corresponding to NUP96) showed elution profiles that clustered together, whereas isoforms 3 and 4 also clustered together, but with a different elution profile than the other isoforms. Fig. 11*B* shows the NUP98 interaction partners as identified by STRING. Fig. 11*C* shows the elution profiles for the NUP96/98 isoforms in our dataset. The profiles for isoforms 3 and 4 (NUP98) are illustrated in orange, and the green profile is that of the four isoforms for NUP96. From these profiles we conclude that protein–protein interactions involving NUP96 occur predominantly in fractions 8–17, whereas those specifically involving the N-terminal region (*i.e.* NUP98) occur predominantly in fractions 22–31.

We interrogated the SEC dataset for elution profiles of known NUP98 interaction partners and illustrate this here for some of these proteins (Fig. 11*D*). These elution profiles show that some reported interaction partners have peaks that overlap with the NUP96 component (*i.e.* SEC13, NUP107, NUP133, NUP160, NUP37, NUP43, and SEH1L), whereas RAE1 has a peak coinciding instead with the N-terminal NUP98 protein peak. This suggests that RAE1 may interact selectively with NUP98, but not with NUP96. These data indicate that the isoform groups corresponding to the respectively named NUP96 and NUP98 proteins are involved in forming distinct protein complexes. Our data also show a bias in detection sensitivity between NUP96 and NUP98 isoforms. We propose that this may have arisen in the native extracts examined because of the different solubility of the NUP98 protein, which is known to be tightly membrane associated, relative to the more soluble NUP96 protein.

Protein phosphorylation is known to affect many protein–protein interactions. The presence of the additional negative charge can increase affinity between interaction partners and thus generate new binding sites, or, alternatively, it can decrease affinity and thereby block interactions (53). We therefore searched our dataset for peptides showing phosphorylation at serine, threonine, or tyrosine residues and used this information to correlate with the elution profiles of the respective modified and unmodified forms of each protein (supplemental Table S3).

NUDT5 is a protein known to form a homodimer (54). It functions in cells to hydrolyze ADP-ribose and other similar molecules, such as 8-oxoGDP, which can be generated by nucleotide-damaging events (55). NUDT5 was detected in this dataset with 13 peptides, representing 64% sequence coverage spanning the full length of the protein (Fig. 12*A*). Using NUDT5 as an example, we observed that its elution profile had two main peaks, one in fraction 21 (molecular weight ~ 440 kDa) and the other in fraction 28 (molecular weight ~ 67 kDa) (Fig. 12*B*).

Our dataset identified a peptide from NUDT5 with phosphorylation of serine 3 that was present in all three biological replicates. Whereas the remaining 12 NUDT5 peptides we identified were present in both peak fractions, including the unphosphorylated form of the serine-3-containing peptide, the serine 3 phosphorylated NUDT5 peptide was exclusively identified in fractions 27–29, corresponding to the second, lower molecular weight peak in the overall protein elution profile (Fig. 12*B*). Interestingly, this site is part of an S-Q phosphorylation motif recognized by the DNA damage responsive kinases ATM and ATR (56). These data suggest that the formation of one or more forms of protein complex containing NUDT5 may be regulated, at least in part, by the phosphorylation of serine 3. These phosphorylation-specific data are represented in the EPD with the detection of phosphopeptides for each protein highlighted on the elution profile of the entire protein based upon nonphosphorylated peptides (Fig. 12*D*).

In summary, the data above illustrate that the combined use of native SEC, MS-based peptide identification, and data analysis can identify candidate protein complexes involving selective interactions of specific protein isoforms and/or post-translationally modified protein forms.

DISCUSSION

In this study we used size exclusion chromatography combined with mass spectrometry in a quantitative proteomics strategy to provide a system-wide analysis of the diverse range of protein complexes formed in human cells. In particular, we have shown how this strategy can be used to annotate proteins with respect to their participation in complexes and to predict the formation of complexes involving specific protein isoforms and post-translationally modified forms. We provide here one of the largest concerted experimental surveys to date of native protein complexes, involving the detection and characterization of the SEC elution profiles for more than 71,600 peptides, corresponding to over 8,000 human proteins, and including the detection of many distinct isoforms and phosphorylation sites. Furthermore, the data are derived from three independent biological replicates, allowing statistical evaluation of the quality of the data during all downstream analyses.

Using hierarchical clustering, we identified proteins that co-elute with known components of protein complexes and thus may represent either previously unidentified interaction partners or entirely new complexes. Using these data, we have demonstrated examples of proteins that form multiple, separable complexes and confirmed these findings in detail for the MRFAP1, MORF4L1, and MRGBP protein interaction network. We have also confirmed, using multiple examples, that the conclusions derived from MS-based protein identifications are supported by parallel antibody-based protein detection using Western blots. To enable easy interrogation of these data and provide wide access to the community, the

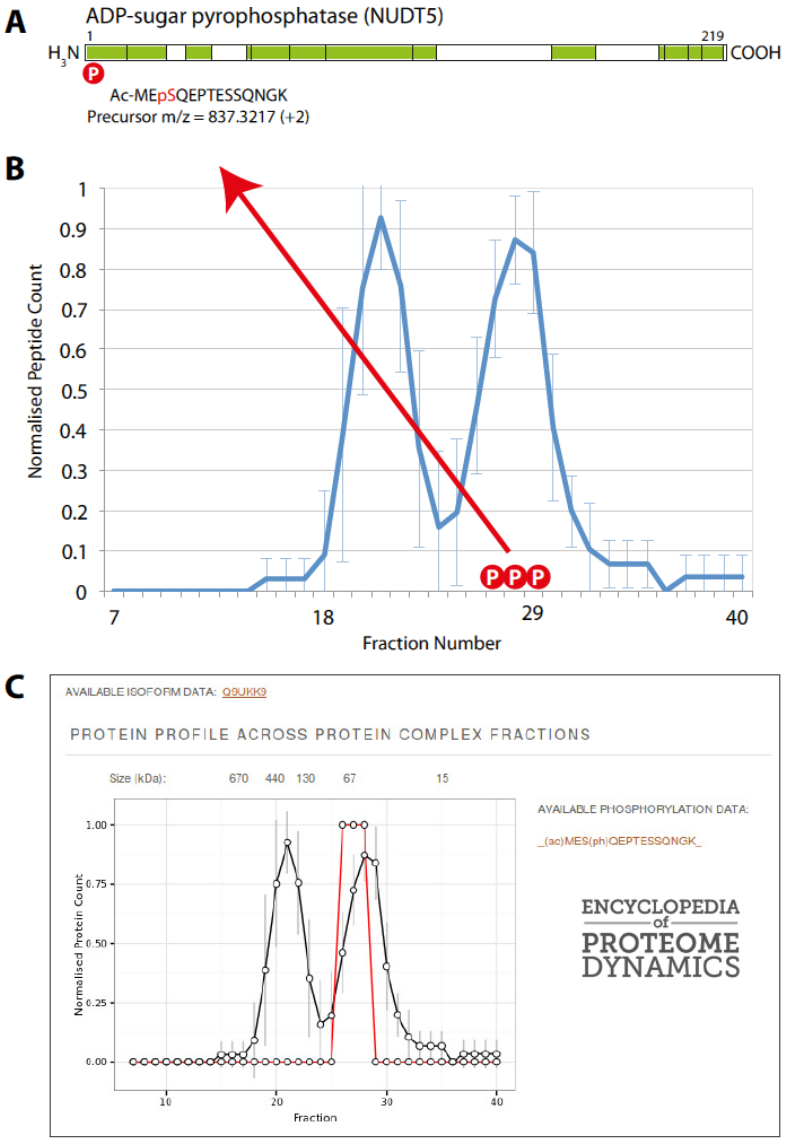


FIG. 12. **Comparison of unmodified and phosphorylated protein profiles.** *A*, graphical representation of the peptides identified for the NUDT5 protein (shown in green). The letter "P" in the red circle denotes a phosphorylation site on the NUDT5 protein. *B*, the line graph illustrates the SEC elution profile for the NUDT5 unmodified protein. Error bars represent the standard deviation from the mean. The letter "P" in a red circle denotes a fraction containing the phosphorylated peptide (Ac-MEpSQEPTESSQNGK). *C*, a screen shot of these data presented in the Encyclopedia of Proteome Dynamics. Elution profiles show the normalized peptide count on the *y*-axis plotted against the SEC fraction on the *x*-axis.

entire dataset has been integrated into our online searchable database, the EPD.

This study demonstrates the utility of the combined SEC-MS approach for the systematic characterization and annotation of protein complexes at a system-wide level. By comparing the results from independent biological replicates, we have shown that the methodology is reproducible and provides deep coverage of cellular protein complexes. For example, analysis of this dataset indicated that we identified in the native extracts from U2OS cells more than 50% of the total protein complexes for humans published in the CORUM database, which is compiled from a cumulative analysis of the

published literature, including data derived from multiple cell types. Importantly, we have also demonstrated that different isoforms of a single protein can be detected, whether they are generated by alternative splicing, proteolytic cleavage, and/or post-translational modifications, and we have provided examples in which such isoforms are involved in distinct complexes and protein–protein interactions that can be differentiated and analyzed systematically. This provides a powerful tool for detecting and annotating protein complexes that augments other techniques commonly used to characterize protein complexes, including affinity pull-down methods and previously described SEC and ion-exchange chromatography methods. Furthermore, by combining these data in the EPD with our previous large-scale measurements of protein properties, such as subcellular localization and protein turnover rates, we have shown here how a Super Experiment approach (17) can be developed to provide a deep, multidimensional annotation of the human proteome.

Although we have demonstrated here the current feasibility of using a combined SEC-MS approach for identifying potential protein complexes and isoform and PTM-dependent interactions, there is considerable room for further technical improvements that will extend the resolution and sensitivity to detect lower abundance proteins and PTMs that are underrepresented using current technology. For example, recent and ongoing developments in mass spectrometry techniques and instrumentation have resulted in large increases in the sensitivity and resolution of proteome analyses, which we took advantage of in this study. Further improvements in both MS technology and the associated software for the analysis of raw MS data will undoubtedly continue and thereby enhance the number of proteins, isoforms, and PTMs that can be routinely detected across all of the SEC fractions. Furthermore, there is a clear opportunity to improve the resolution and size range of protein complex fractionation through future advances in SEC column technology and the advent of advanced ultra-high-performance liquid chromatography systems. Considering that even with the limitations of the resolution currently available with SEC, it is still possible to separate protein complexes at a level that is sufficient to derive a large amount of useful biological information, future improvements in SEC technology should make this approach substantially more powerful. For example, we anticipate that modifications in the design of SEC columns and improvements in silica matrices will result in improved resolution that will allow better separation of protein complexes such as the chaperonin containing TCP1 and elongation factor complexes shown in Fig. 6*D*. Potentially, this could also allow the effective separation of complexes up to the largest protein complexes present in cells, such as the nuclear pore complex at 45 MDa. Improvements in the algorithms used for the clustering of proteins and the *in silico* detection of their component peaks will also allow the identification of more potential protein–protein interactions, particularly for proteins present in more than one independent complex.

We and others have previously combined mass-spectrometry-based proteomics with cell fractionation to measure the subcellular localization of the proteome. For example, we have recently examined differences in protein isoform distributions and properties such as degradation rates between subcellular compartments and have evaluated how such properties are changed upon cell perturbations (39). Combining the SEC methodology with cellular fractionation strategies, drug treatments, and analyses of specific cell cycle stages will allow detailed insight into how the dynamics of protein complex formation is controlled in space and time. For example, in a recent study it has been shown that protein complexes in cytoplasmic extracts can be separated and characterized via SEC combined with mass spectrometry (21). This study combined the SEC approach with stable isotope labeling of amino acids in cell culture to analyze temporal protein complex changes following epidermal growth factor stimulation in cytoplasmic extracts of HeLa cells. Analysis of complexes can also be applied at the organelle level, using methods described for the proteomic analysis of organelles fractionated by means of sucrose gradients and/or differential centrifugation (57), prior to SEC and subsequent LC-MS/MS analysis.

When the different methods for studying protein complexes are compared, an advantage of SEC that is apparent is the ability to compare in parallel the complexes that are recovered using a variety of different buffer conditions. We elected to use a buffer containing no detergent and physiological levels of sodium chloride to maintain the integrity of native protein complexes. However, the buffer can be varied systematically, for example, to increase the salt concentration and/or include increasing levels of detergents, which will reveal the differential stabilities of different forms of complexes. In addition, the use of detergents to solubilize various membrane-associated complexes and protein subdomains could also be informative and may allow the recovery and identification of a greater proportion of the human proteome than is possible using native extraction conditions. In this regard it will also be interesting to explore the use of cross-linking agents in conjunction with SEC fractionation to improve the coverage of lower affinity complexes. The use of cross-linkers has recently been shown to improve the coverage of protein complex components in IP analyses (58).

Compared with other techniques for protein separation, SEC has the advantage of providing a relatively high capacity for total protein loading. In addition, as the proteins are in solution throughout the separation, the yield and efficiency of proteomics workflows is improved, and the need to perform in-gel enzyme digestions and subsequent ex-gel peptide isolation is avoided. Havugimana *et al.* used a combination of ion-exchange chromatography with four analytical column combinations, sucrose gradient centrifugation, and isoelectric

focusing to analyze soluble protein complexes formed by 5,584 proteins in HeLa S3 and HEK293 cells (20). By combining SEC with one or more of these other forms of HPLC chromatography, the separation of different forms of complexes could be enhanced further. However, not all complexes will remain stable using methods that, for example, require higher salt treatments as part of the elution or separation procedure. An advantage of SEC is that separation can be performed throughout in defined native buffer conditions.

We are optimistic that future applications of the SEC technique will provide a framework for the system-wide comparison of protein complexes under a wide range of conditions and responses. This could include the analysis of complexes at different cell cycle stages or following cellular perturbations, including drug treatments, as well as comparisons of differences in protein complexes between cell and tissue types. Such systematic, high-throughput, system-wide comparisons of protein complexes under different conditions are simply not feasible with the current affinity-purification-based methods, not least because of the sheer complexity and associated cost of performing and analyzing many thousands of pull-downs under multiple experimental conditions. Another limitation is associated with overexpression of the tagged bait and the fact that such exogenous proteins are often co-expressed with the endogenous, untagged protein and have to compete for binding to partner proteins and for incorporation into complexes. This is particularly relevant in human cells and in model organisms in which convenient genetic strategies are not available for direct replacement at the original locus of the endogenous gene with a transgene encoding the tagged bait. We note that the SEC approach directly analyzes untagged, endogenous protein complexes and thus avoids the need to construct thousands of cell lines expressing tagged constructs in order to compare large numbers of cellular complexes.

Size exclusion chromatography allows the separation of individual proteins that participate in multiple complexes of different sizes and/or shapes. In this work we have highlighted analysis of the MORF4L1 adaptor protein, which is an example of a protein that is a component of several large complexes, including the NuA4 (59), Sin3A (44), and BRCA1-PALB2 complexes (60). The MRFAP1 protein was known to interact with MORF4L1 (61), but the resulting complex had not been characterized. In our SEC dataset, the elution profiles of the MORF4L1 and MRFAP1 proteins were seen to overlap only in the small size fractions. This showed that the interaction occurred in a small complex that was separable and clearly distinct from the larger MORF4L1-containing complexes mentioned above. We note that when analyzed via immunoaffinity purification, all of the MORF4L1 complexes and interaction partners co-purified together. In this case the use of SEC provided the size resolution and prefractionation of the separate forms of complexes necessary to help delineate differences in their protein compositions.

Through SEC analysis, three different types of protein isoforms were found to form isoform-specific interactions with distinct protein complexes. This was typified by the NUDT5 protein, which is a nucleotide pyrophosphatase that is known to hydrolyze a variety of substrates, including ADP-ribose, ADP-mannose, ADP-glucose, 8-oxo-GDP, and 8-oxo-dGDP. Interestingly, the substrates 8-oxo-dGDP and 8-oxo-GDP are produced following the actions of reactive oxygen species on nucleotides, and this can lead to either mutations in the genome or mistranslation of mRNA (62).

NUDT5 is a ~25-kDa protein that is known to form a homodimeric structure. This corresponds with the estimated size of the NUDT5 peak with an elution maximum in fraction 28 from our SEC data, and this was the same peak where we detected NUDT5 to be phosphorylated at serine 3. This phosphosite in NUDT5 has also been identified previously. We note that this phosphorylation site on NUDT5 was identified as an ATM/ATR kinase substrate motif in a large-scale study using Phospho-ATM/ATR substrate (S*Q) immunoaffinity beads (56). Therefore, we propose that phosphorylation at this site may activate NUDT5 to hydrolyze damaged nucleotides. We also hypothesize that the separate peak, with maximum elution in fraction 20, corresponds to a form of NUDT5 that was further multimerized and/or interacting with additional partner proteins.

These data on NUDT5 serve to illustrate the potential for mining complex data within the EPD to generate predictions that can be used to design further experiments and investigate potential new regulatory mechanisms. We note that the analysis of the SEC data also underlines the importance of achieving high peptide sequence coverage when identifying proteins. High sequence coverage enhances our ability to discriminate between separate protein isoforms (63). This in turn allows us to identify examples of protein complexes that involve isoform-specific interactions, as illustrated here for HP1BP3.

An unexpected result of this study was the observation that there were systematic differences in the likely biological functions performed by proteins in different-sized complexes, as judged by the frequency of association of GO terms with proteins in either the larger or smaller classes of complexes. The finding that smaller complexes are generally enriched for terms associated with metabolic processes may signify that these proteins/complexes generally require fewer subunits for enzymatic function and regulation. For example, many metabolic reactions may generate a single product from a specific substrate, with the modification being made to the substrate often by one enzyme, albeit with multiple subunits. The resulting product can move to the next enzyme in the pathway by diffusion, without a requirement for any additional protein–protein interactions. In contrast, mechanistically complex biological processes, such as vesicle transport and steps in gene expression, may require many proteins to work together within large, integrated complexes, such as, for example, RNA

polymerases and the exocyst and COG complexes. Other large complexes, such as the CUL3 E3 ligase complex, may also rely on larger numbers of protein components to mediate regulation of their function in response to temporal and/or spatial cues. In contrast, smaller enzymatic complexes involved in intermediary metabolism, such as dimeric GAPDH, can have their activity modulated by allosteric regulation. This may require fewer, if any, additional protein–protein interactions to facilitate substrate recognition or regulatory functions and hence integrate their activity with metabolic homeostasis.

In summary, this study provides a resource identifying soluble, native protein complexes in U2OS cells, separated on the basis of size and shape. This work incorporates data on protein isoforms and phosphorylation sites and statistical evaluation of independent biological replicates. We have grouped proteins with similar elution profiles into clusters, thereby identifying both well-defined and known complexes and novel potential protein interaction networks. The workflow described here can be used and extended in the future to aid the system-wide characterization of protein complexes and their dynamics in cells and tissues. The resulting information will be provided for the community in the online Encyclopedia of Proteome Dynamics, thereby enhancing the value of the protein complex data by making it easy to search, visualize, and analyze.

Acknowledgments—We thank Dr. Eric Griffis for useful discussions. K.K. is a Wellcome Trust Clinical Ph.D. Student. A.I.L. is a Wellcome Trust Principal Research Fellow.

* This work was supported by grants from the Wellcome Trust (Grant Nos. 083524/Z/07/Z, 097945/B/11/Z, 073980/Z/03/Z, 08136/Z/03/Z, and 0909444/Z/09/Z), the EU FP7 Prospects network (Grant No. HEALTH-F4-2008-201648), EpiGeneSys network (Grant No. HEALTH-F4-2010-257082), and the BBSRC LoLa (Grant No. BB/K003801/1).

Ⓢ This article contains supplemental material.

¶ To whom correspondence should be addressed: Angus I. Lamond, Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow St., Dundee, United Kingdom, Tel.: 44-1382385473, Fax: 44-1382385695.

§ These authors contributed to this work equally.

REFERENCES

1. Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.-W. (2010) CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* **38,** D497–D501
2. Dunham, W. H., Mullin, M., and Gingras, A.-C. (2012) Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **12,** 1576–1590
3. Malovannaya, A., Li, Y., Bulynko, Y., Jung, S. Y., Wang, Y., Lanz, R. B., O'Malley, B. W., and Qin, J. (2010) Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 2431–2436
4. Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulynko, Y., Le, N. T., Chan, D. W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., Kim, B.-J., Li, C., Chen, R., Li, W., Wang, Y., O'Malley, B. W., and Qin, J. (2011) Analysis of the human endogenous coregulator complexome. *Cell* **145,** 787–799
5. Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y. V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J.-P., Duewel, H. S., Stewart, I. I., Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S.-L., Moran, M. F., Morin, G. B., Topaloglou, T., and Figeys, D. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3,** 89
6. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W. V., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415,** 180–183
7. Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.-A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.-M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636
8. Collins, M. O., and Choudhary, J. S. (2008) Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotech.* **19,** 324–330
9. Gingras, A.-C., Gstaiger, M., Raught, B., and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **8,** 645–654
10. Goel, A., Colcher, D., Koo, J.-S., Booth, B. J. M., Pavlinkova, G., and Batra, S. K. (2000) Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *Biochim. Biophys. Acta* **1523,** 13–20
11. Rumlová, M., Benedíková, J., Cubínková, R., Pichová, I., and Ruml, T. (2001) Comparison of classical and affinity purification techniques of Mason-Pfizer monkey virus capsid protein: the alteration of the product by an affinity tag. *Protein Expr. Purif.* **23,** 75–83
12. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386
13. Ong, S.-E., and Mann, M. (2007) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1,** 2650–2660
14. Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189,** 739–754
15. Trinkle-Mulcahy, L., Boulon, S., Lam, Y. W., Urcia, R., Boisvert, F.-M., Vandermoere, F., Morrice, N. A., Swift, S., Rothbauer, U., Leonhardt, H., and Lamond, A. (2008) Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* **183,** 223–239
16. Trinkle-Mulcahy, L., Andersen, J., Lam, Y. W., Moorhead, G., Mann, M., and Lamond, A. I. (2006) Repo-Man recruits PP1γ to chromatin and is essential for cell viability. *J. Cell Biol.* **172,** 679–692
17. Boulon, S., Ahmad, Y., Trinkle-Mulcahy, L., Verheggen, C., Cobley, A., Gregor, P., Bertrand, E., Whitehorn, M., and Lamond, A. I. (2010) Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners. *Mol. Cell. Proteomics* **9,** 861–879
18. Heide, H., Bleier, L., Steger, M., Ackermann, J., Dröse, S., Schwamb, B., Zörnig, M., Reichert, A. S., Koch, I., Wittig, I., and Brandt, U. (2012) Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell Metab.* **16,** 538–549
19. Camacho-Carvajal, M. M., Wollscheid, B., Aebersold, R., Steimle, V., and Schamel, W. W. A. (2004) Two-dimensional blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. *Mol. Cell. Proteomics* **3,** 176–182
20. Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z.,

Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V.-u.-N., Bezginov, A., Clark, G. W., Wu, G. C., Wodak, S. J., Tillier, E. R. M., Paccanaro, A., Marcotte, E. M., and Emili, A. (2012) A census of human soluble protein complexes. *Cell* **150,** 1068–1081

21. Kristensen, A. R., Gsponer, J., and Foster, L. J. (2012) A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* **9,** 907–909
22. Olinares, P. D. B., Ponnala, L., and van Wijk, K. J. (2010) Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **9,** 1594–1615
23. Wheaton, R. M., and Bauman, W. C. (1953) Non-ionic separations with ion exchange resins. *Ann. N. Y. Acad. Sci.* **57,** 159–176
24. Wessel, D., and Flügge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138,** 141–143
25. Haas, W., Faherty, B. K., Gerber, S. A., Elias, J. E., Beausoleil, S. A., Bakalarski, C. E., Li, X., Villén, J., and Gygi, S. P. (2006) Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **5,** 1326–1337
26. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372
27. Cox, J. r., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805
28. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57
29. Consortium, T. U. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **41,** D43–D47
30. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7,** 548
31. Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **11,** M111.014050
32. Snel, B., Lehmann, G., Bork, P., and Huynen, M. A. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28,** 3442–3444
33. Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L. J., and Mering, C. v. (2010) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39,** D561–D568
34. Leitner, A., Joachimiak, L. A., Bracher, A., Mönkemeyer, L., Walzthoeni, T., Chen, B., Pechmann, S., Holmes, S., Cong, Y., Ma, B., Ludtke, S., Chiu, W., Hartl, F. U., Aebersold, R., and Frydman, J. (2012) The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20,** 814–825
35. Verschoor, A., Srivastava, S., Grassucci, R., and Frank, J. (1996) Native 3D structure of eukaryotic 80s ribosome: morphological homology with E. coli 70S ribosome. *J. Cell Biol.* **133,** 495–505
36. Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J., and Robinson, C. V. (2006) Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol.* **4,** e267
37. Coll, J. M., Sekowski, J. W., Hickey, R. J., Schnaper, L., Yue, W., Brodie, A. M., Uitto, L., Syvaoja, J. E., and Malkas, L. H. (1996) The human breast cell DNA synthesome: its purification from tumor tissue and cell culture. *Oncol. Res.* **8,** 435–447
38. Coll, J. M., Hickey, R. J., Cronkey, E. A., Jiang, H. Y., Schnaper, L., Lee, M. Y., Uitto, L., Syvaoja, J. E., and Malkas, L. H. (1997) Mapping specific protein-protein interactions within the core component of the breast cell DNA synthesome. *Oncol. Res.* **9,** 629–639
39. Larance, M., Ahmad, Y., Kirkwood, K. J., Ly, T., and Lamond, A. I. (2013) Global subcellular characterization of protein degradation using quantitative proteomics. *Mol. Cell. Proteomics* **12,** 638–650
40. Uhlén, M., Björling, E., Agaton, C., Szigyarto, C. A.-K., Amini, B., Andersen, E., Andersson, A.-C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Ödling, J., Oksvold, P., Olsson, I., Öster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, Å., Sköllermo, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4,** 1920–1932
41. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28,** 1248–1250
42. Cai, Y., Jin, J., Tomomori-Sato, C., Sato, S., Sorokina, I., Parmely, T. J., Conaway, R. C., and Conaway, J. W. (2003) Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. *J. Biol. Chem.* **278,** 42733–42736
43. Jelinic, P., Pellegrino, J., and David, G. (2011) A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol. Cell. Biol.* **31,** 54–62
44. Yochum, G. S., and Ayer, D. E. (2002) Role for the mortality factors MORF4, MRGX, and MRG15 in transcriptional repression via associations with Pf1, mSin3A, and transducin-like enhancer of split. *Mol. Cell. Biol.* **22,** 7868–7876
45. Xie, T., Graveline, R., Kumar, G. S., Zhang, Y., Krishnan, A., David, G., and Radhakrishnan, I. (2012) Structural basis for molecular interactions involving MRG domains: implications in chromatin biology. *Structure* **20,** 151–160
46. Kumar, G. S., Xie, T., Zhang, Y., and Radhakrishnan, I. (2011) Solution structure of the mSin3A PAH2–Pf1 SID1 complex: a Mad1/Mxd1-like interaction disrupted by MRG15 in the Rpd3S/Sin3S complex. *J. Mol. Biol.* **408,** 987–1000
47. Leung, J. K., Berube, N., Venable, S., Ahmed, S., Timchenko, N., and Pereira-Smith, O. M. (2001) MRG15 activates the B-myb promoter through formation of a nuclear complex with the retinoblastoma protein and the novel protein PAM14. *J. Biol. Chem.* **276,** 39171–39178
48. Zhang, P., Zhao, J., Wang, B., Du, J., Lu, Y., Chen, J., and Ding, J. (2006) The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. *Protein Sci.* **15,** 2423–2434
49. Allard, S., Utley, R. T., Savard, J., Clarke, A., Grant, P., Brandl, C. J., Pillus, L., Workman, J. L., and Cote, J. (1999) NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J.* **18,** 5108–5119
50. Larance, M., Kirkwood, K. J., Xirodimas, D. P., Lundberg, E., Uhlen, M., and Lamond, A. I. (2012) Characterization of MRFAP1 turnover and interactions downstream of the NEDD8 pathway. *Mol. Cell. Proteomics* **11,** M111.014407
51. Hayashihara, K., Uchiyama, S., Shimamoto, S., Kobayashi, S., Tomschik, M., Wakamatsu, H., No, D., Sugahara, H., Hori, N., Noda, M., Ohkubo, T., Zlatanova, J., Matsunaga, S., and Fukui, K. (2010) The middle region of an HP1-binding protein, HP1-BP74, associates with linker DNA at the entry/exit site of nucleosomal DNA. *J. Biol. Chem.* **285,** 6498–6507
52. Fontoura, B. M. A., Blobel, G., and Matunis, M. J. (1999) A conserved biogenesis pathway for nucleoporins: proteolytic processing of a 186-kilodalton precursor cenerates Nup98 and the novel nucleoporin, Nup96. *J. Cell Biol.* **144,** 1097–1112
53. Nishi, H., Hashimoto, K., and Panchenko, A. R. (2011) Phosphorylation in protein-protein binding: effect on stability and function. *Structure* **19,** 1807–1815
54. Zha, M., Zhong, C., Peng, Y., Hu, H., and Ding, J. (2006) Crystal structures of human NUDT5 reveal insights into the structural basis of the substrate specificity. *J. Mol. Biol.* **364,** 1021–1033
55. Ishibashi, T., Hayakawa, H., and Sekiguchi, M. (2003) A novel mechanism for preventing mutations caused by oxidation of guanine nucleotides. *EMBO Rep.* **4,** 479–483
56. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., and Zhang, B.

(2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4,** 1551–1561

57. Gatto, L., Vizcaíno, J. A., Hermjakob, H., Huber, W., and Lilley, K. S. (2010) Organelle proteomics experimental designs and analysis. *Proteomics* **10,** 3957–3969
58. Humphries, J. D., Byron, A., Bass, M. D., Craig, S. E., Pinney, J. W., Knight, D., and Humphries, M. J. (2009) Proteomic analysis of integrin-associated complexes identifies RCC2 as a dual regulator of Rac1 and Arf6. *Sci. Signal.* **2,** ra51
59. Doyon, Y., and Cote, J. (2004) The highly conserved and multifunctional NuA4 HAT complex. *Curr. Opin. Genet. Dev.* **14,** 147–154
60. Hayakawa, T., Zhang, F., Hayakawa, N., Ohtani, Y., Shinmyozu, K., Nakayama, J.-i., and Andreassen, P. R. (2010) MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. *J. Cell Sci.* **123,** 1124–1130
61. Tominaga, K., Magee, D. M., Matzuk, M. M., and Pereira-Smith, O. M. (2004) PAM14, a novel MRG- and Rb-associated protein, is not required for development and T-cell function in mice. *Mol. Cell. Biol.* **24,** 8366–8373
62. Sekiguchi, T., Ito, R., Hayakawa, H., and Sekiguchi, M. (2013) Elimination and utilization of oxidized guanine nucleotides in the synthesis of RNA and its precursors. *J. Biol. Chem.* **288,** 8128–8135
63. Ahmad, Y., Boisvert, F.-M., Lundberg, E., Uhlen, M., and Lamond, A. I. (2012) Systematic analysis of protein pools, isoforms, and modifications affecting turnover and subcellular localization. *Mol. Cell. Proteomics* **11,** M111.013680

Research

⌘ *Author's Choice*

Characterization of MRFAP1 Turnover and Interactions Downstream of the NEDD8 Pathway*[S]

Mark Larance‡, Kathryn J. Kirkwood‡, Dimitris P. Xirodimas‡§, Emma Lundberg¶, Mathias Uhlen¶, and Angus I. Lamond‡‖

The NEDD8-Cullin E3 ligase pathway plays an important role in protein homeostasis, in particular the degradation of cell cycle regulators and transcriptional control networks. To characterize NEDD8-cullin target proteins, we performed a quantitative proteomic analysis of cells treated with MLN4924, a small molecule inhibitor of the NEDD8 conjugation pathway. MRFAP1 and its interaction partner, MORF4L1, were among the most up-regulated proteins after NEDD8 inhibition in multiple human cell lines. We show that MRFAP1 has a fast turnover rate in the absence of MLN4924 and is degraded via the ubiquitin-proteasome system. The increased abundance of MRFAP1 after MLN4924 treatment results from a decreased rate of degradation. Characterization of the binding partners of both MRFAP1 and MORF4L1 revealed a complex protein-protein interaction network. MRFAP1 bound to a number of E3 ubiquitin ligases, including CUL4B, but not to components of the NuA4 complex, including MRGBP, which bound to MORF4L1. These data indicate that MRFAP1 may regulate the ability of MORF4L1 to interact with chromatin-modifying enzymes by binding to MORF4L1 in a mutually exclusive manner with MRGBP. Analysis of MRFAP1 expression in human tissues by immunostaining with a MRFAP1-specific antibody revealed that it was detectable in only a small number of tissues, in particular testis and brain. Strikingly, analysis of the seminiferous tubules of the testis showed the highest nuclear staining in the spermatogonia and much weaker staining in the spermatocytes and spermatids. MRGBP was inversely correlated with MRFAP1 expression in these cell types, consistent with an exchange of MORF4L1 interaction partners as cells progress through meiosis in the testis. These data highlight an important new arm of the NEDD8-cullin pathway. ***Molecular & Cellular Proteomics 11: 10.1074/mcp.M111.014407, 1–11, 2012.***

Quantitative proteomic analysis has become the gold standard in recent years for unbiased analysis of cellular responses to drug treatment and the analysis of protein-protein interactions. Stable isotope labeling with amino acids in cell culture (SILAC)[1] is one of the most accurate techniques for performing mass spectrometry-based quantitative proteomics experiments (1). SILAC has been used to study a wide range of cellular protein responses, including the reliable detection of specific protein interaction partners, subcellular protein localization, and changes in protein levels resulting from viral infection, drug treatments, and stress responses (2). Pulsed incorporation of amino acids containing stable isotopes was also used to measure protein turnover in nucleoli (3). Recently, detailed proteome-wide studies have used pulsed SILAC to calculate the synthesis and degradation rates of human and mouse proteins, which gives insight into protein regulation and dynamics (4, 5). In both of these studies, a subset of proteins was identified with short half-lives (<3 h). Among this group are two known interaction partners, *i.e.* a chromatin regulatory protein called mortality factor 4-like protein 1 (MORF4L1 or MRG15) and MORF4 family-associated protein 1 (MRFAP1).

Protein turnover plays an essential role in regulating cellular proliferation and differentiation. The ubiquitination of protein substrates mediates a large proportion of protein degradation within the cell. There are more than 600 E3 ubiquitin ligases encoded by the human genome, and of these the Cullin-based E3 ligase complexes are the most numerous and complex. The activity of Cullin E3 ligase complexes is controlled in part by their post-translational modification by the ubiquitin-like peptide NEDD8 (6, 7). NEDDylation is performed in an analogous way to ubiquitination, with NEDD8 initially binding

From the ‡Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee, DD1 5EH, United Kingdom and the ¶Science for Life Laboratory, School of Biotechnology, KTH, Tomtebodavägen 23A, SE-171 65 Solna, Stockholm, Sweden

⌘ *Author's Choice*—Final version full access.

Received September 15, 2011, and in revised form, October 28, 2011

Published, MCP Papers in Press, October 29, 2011, DOI 10.1074/mcp.M111.014407

[1] The abbreviations used are: SILAC, stable isotope labeling with amino acids in cell culture; CUL4B, Cullin-4b; MORF4, mortality factor 4; MORF4L1, mortality factor 4-like protein 1; MRFAP1, MORF4 family-associated protein 1; MRGBP, MRG domain-binding protein; mRNA, messenger RNA; NEDD8, neural precursor cell expressed developmentally down-regulated protein 8; NuA4, nucleosome acetyltransferase of histone H4; TBST, TBS Tween 20; E1, ubiquitin-activating enzyme; E2, ubiquitin carrier protein; E3, ubiquitin-protein isopeptide ligase; MES, 4-morpholineethanesulfonic acid; IP, immunoprecipitation.

an E1 enzyme complex (APP-BP1/Uba3) (8), followed by two possible E2 enzymes for conjugation to substrates (8, 9). Several enzymes are able to catalyze deNEDDylation, at least *in vitro*, with the most well studied being the COP9 signalosome (10).

Previous studies have shown that NEDD8 conjugation to Cullins increases their ubiquitination activity and thus the degradation rate for their substrate proteins (11). A specific inhibitor of the NEDD8 E1 enzyme complex has been identified, called MLN4924 (12). MLN4924 is currently in clinical trials as an anti-cancer agent because of its ability to trigger cellular senescence, which is thought to occur after uncontrolled DNA rereplication in the S phase (12, 13). MLN4924 rapidly blocks NEDDylation of substrate proteins, including the Cullins. A recent study has reported a large scale proteomic and transcriptomic analysis of genes whose expression is altered upon blocking NEDD8 conjugation with MLN4924 in A375 human melanoma cells (14). This study identified 38 proteins as crucial to the cytotoxicity of MLN4924 in pathways such as cell cycle regulation, DNA repair, and ubiquitination (14).

The human NuA4 histone acetyltransferase complex is highly conserved from yeast to humans and is critical for the acetylation of histone H4 to regulate chromatin activity (15). Two conserved components of the NuA4 complex are MORF4L1 and MRG-binding protein (MRGBP) (15). MORF4L1 contains two domains separated by a linker region, the first being an N-terminal chromo domain that has been shown to have affinity for di- or tri-methylated histone H3 on K36 (16, 17). The second domain is the C-terminal MRG domain (MORF4-related gene domain), which has been shown to interact with MRGBP and MRFAP1 (17, 18, 20). The interaction of MORFL41 with the MRGBP protein is thought to recruit the NuA4 complex to chromatin via its chromo domain (15). The exact function of the interaction between MORF4L1 and MRFAP1 is unknown (21). However, we note that these interacting proteins were among the most rapidly turned over proteins identified in mammalian cells in the recent pulse-SILAC studies (4, 5).

In the present study, we have taken advantage of the specificity of the NEDD8 inhibitor MLN4924, combined with SILAC-based quantitative mass spectrometry methods to characterize proteins downstream of the NEDD8 pathway. Here, we report that MRFAP1, a component of the MORF4L1 complex, was stabilized dramatically by MLN4924 treatment. We present evidence based upon both protein-protein interaction studies and analysis of expression in human tissues suggesting a model in which MRFAP1 competes with MRGBP for binding to MORF4L1, thereby regulating MORF4L1 function in chromatin modification. Furthermore, examination of the human tissue expression of MRFAP1 showed that it may play a critical role in spermatogenesis, possibly by regulating the hyperacetylation of chromatin on histone H4.

EXPERIMENTAL PROCEDURES

Materials—U2OS, U138MG, Tera-1, and HaCaT cells were purchased from the American Type Culture Collection (Rockville, MD). Dulbecco's modified Eagle's medium, minimal essential medium, McCoy's 5a medium, fetal calf serum, antibiotics, NuPage gels, LDS sample buffer, MES SDS-PAGE running buffer, nitrocellulose iBlot stacks, SYPRO Ruby, Alexa Fluor 680-conjugated secondary antibodies, and CBQCA assay kit were obtained from Invitrogen. IrDye 800-conjugated secondary antibodies were obtained from Rockland Immunochemicals (Gilbertsville, PA). Histo-Clear was from National Diagnostics (Atlanta, GA). antigen unmasking solution (citric acid based) was from Vector Laboratories (Burlingame, CA). Horseradish peroxidase-conjugated secondary antibodies were from Cell Signaling Technology (Danvers, MA). Bicinchoninic acid assay reagents, Coomassie Plus (Bradford) reagent, a subcellular protein fractionation kit, detergent removal plates, 16% methanol-free paraformaldehyde, and triscarboxyethylphosphine (Bond-breaker neutral pH solution) were from Pierce. Trypsin Gold was from Promega. Sep-Pak tC18 96-well μ-elution plates were from Waters. The Pepmap C18 columns and trapping cartridges were from Dionex. Complete protease inhibitor mixture tablets and PhosStop phosphatase inhibitor tablets were from Roche Applied Science. All other materials were obtained from Sigma.

Construction of LAP1-MRFAP1 Constructs and Cell Lines—All MRFAP1 constructs were generated by gene synthesis and cloned into the pgLAP1 vector (protein sequences are provided in supplemental Fig. 3) using Gateway cloning (Invitrogen) as described previously (22).

Cell Culture—Briefly, U2OS cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% FCS, 2 mM L-glutamine, 100 units/liter penicillin, and 100 μg/liter streptomycin at 37 °C in 10% CO_2 and passaged at ~80% confluence. U2OS cells expressing LAP1-tagged proteins were grown in the same medium as U2OS but with the addition of 150 μg/ml hygromycin B and 15 μg/ml blasticidine HCl. HaCaT cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% FCS, 100 units/liter penicillin, and 100 μg/liter streptomycin at 37 °C in 10% CO_2 and passaged at ~80% confluence. Tera-1 cells were cultured in McCoy's 5a medium supplemented with 20% FCS, 100 units/liter penicillin, and 100 μg/liter streptomycin at 37 °C in 10% CO_2 and passaged at ~80% confluence. U138MG cells were cultured in minimal essential medium supplemented with 10% FCS, 100 units/liter penicillin, and 100 μg/liter streptomycin at 37 °C in 10% CO_2, and passaged at ~80% confluence. For SILAC labeling of U2OS cells, arginine- and lysine-free Dulbecco's modified Eagle's medium was used and supplemented with stable isotope-labeled arginine and lysine in addition to dialyzed FCS as described previously (23).

Subcellular Fractionation of U2OS Cells—For the initial SILAC screen U2OS cells were treated with either DMSO or MLN4924, combined in a 1:1 ratio of cells and fractionated using differential centrifugation and sucrose cushions as described previously to obtain crude cytoplasmic, nucleoplasmic, and nucleolar fractions (24). These three fractions were then chloroform methanol-precipitated (25) and further separated by molecular weight using denaturing size exclusion chromatography prior to digestion and LC-MS/MS analysis.

A Pierce subcellular protein fractionation kit was used to fractionate U2OS cells according to the manufacturer's instructions. Each fraction was chloroform methanol-precipitated prior to analysis by SDS-PAGE and immunoblotting.

SDS-PAGE and Immunoblotting—Equal amounts of protein were loaded for SDS-PAGE of each sample with 10 μg/lane, except for immunoprecipitation eluates where 10% of the elution volume was loaded. SDS-PAGE was performed using on 4–12% (w/v) Bis-Tris NuPage gels using MES running buffer according to manufacturer's

instructions but with the addition of 25 mM triscarboxyethylphosphine, in the LDS sample buffer. Equal amounts of protein were loaded with a maximum of 10 μg/lane. SYPRO Ruby staining was performed as per manufacturer's instructions (Invitrogen). For mass spectrometric identification SYPRO Ruby staining was performed as per the manufacturer's instructions. For Western blotting, separated proteins were electrophoretically transferred to an iBlot nitrocellulose membrane, blocked with 3% nonfat skim milk in 0.1% Tween 20 in TBS (TBST), and incubated with primary antibody in 5% BSA in TBST overnight at 4 °C. After incubation, the membranes were washed three times in TBST and incubated with horseradish peroxidase-labeled or Alexa fluor 680/IrDye 800-labeled secondary antibodies in 3% nonfat skim milk in TBST. Proteins were visualized using Immobilon chemiluminescent substrate (Millipore) and imaged with a cooled CCD camera (Fuji) for horseradish peroxidase-labeled secondary antibodies or a Licor Odyssey imager for Alexa Fluor 680/IrDye 800-labeled secondary antibodies.

Denaturing Gel Filtration Chromatography, Trypsin Digestion, and Peptide Clean-up—Using a Dionex Ultimate 3000 HPLC system, fractions resuspended in 4% SDS, 100 mM NaCl, 25 mM triscarboxyethylphosphine, 50 mM *N*-ethylmaleimide, 10 mM Na PO_4, pH 6.0, were heated to 65 °C for 10 min and then filtered through a 0.45-μm filter. The samples were injected (20 μl/injection for 100 μg of protein) onto a mAbPacSEC column (Dionex) equilibrated with 0.2% SDS, 100 mM NaCl, 10 mM Na PO_4, pH 6.0. The flow rate was 0.2 ml min^{-1}, and 16 fractions of 100 μl were collected using a low protein binding 96-deep well plate (Eppendorf). Triethylamine bicarbonate (1 M, pH 8.0) was added to each fraction to adjust the pH to 8.0, and trypsin diluted in 0.1 M triethylamine bicarbonate was added at a ratio of 1:50 with incubation for 18 h at 37 °C. SDS was removed from each fraction using detergent removal resin in 96-well plates as described previously (26). For peptide desalting, trifluoroacetic acid was added to 1% (v/v) final concentration, and peptides were purified using a Sep-Pak tC18 96-well μ-elution plate. The peptides were eluted in 200 μl of 50% (v/v) acetonitrile and dried with a SpeediVac prior to resuspension in 5% (v/v) formic acid. Peptide concentrations were determined using the CBQCA assay after 25-fold dilution of peptide samples in 0.1 M borate buffer, pH 9.3.

LC-MS/MS and Maxquant Analysis—Using a Dionex Ultimate 3000 nanoHPLC system, 1 μg of peptides in 5% (v/v) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Dionex). After washing with 2% (v/v) acetonitrile, 0.1% (v/v) formic acid peptides were resolved on a 150-mm × 75-μm Acclaim PepMap C18 reverse phase analytical column over a 100-min organic gradient with a flow rate of 300 nl min^{-1}. The peptides were ionized by nano-electrospray ionization at 1.2 kV using a fused silica emitter with an internal diameter of 5 μm (New Objective). Tandem mass spectrometry analysis was carried out on a LTQ-Velos Orbitrap mass spectrometer Thermo Scientific. The data-dependent acquisition method used was the FT10 protocol as described previously (27). The data were processed, searched, and quantified using the Maxquant software package version 1.2.0.18 as described previously (28), using the default settings and employing the human Uniprot database (June 7, 2011) containing 109,824 entries. The settings used for the Maxquant analysis were: two failed cleavages were allowed; fixed modification was *N*-ethylmaleimide on cysteine; enzymes were trypsin (K/R not before P); variable modifications included in the analysis were methionine oxidation, deamidation of glutamine or asparagine, N-terminal pyroglutamic acid formation, and protein N-terminal acetylation. A mass tolerance of 7 ppm was used for precursor ions, and a tolerance of 0.5 Da was used for fragment ions. Using the default Maxquant settings, a maximum false positive rate of 1% was allowed for both peptide and protein identification. This cut-off was used for accepting individual spectra, as well as whole proteins in the Maxquant output. This threshold has previously been shown to be a rigorous method for identifying true positive matches (28). All of the replicates indicated are biological replicates. Protein quantitation data were always derived from a minimum of two or more peptides/protein.

GFP-IP from SILAC-labeled U2OS Cells—The cells for each condition were harvested separately by trypsinization, washed in PBS, and lysed in IP buffer (1% Nonidet P-40, 50 mM Tris-Cl, pH 7.4, 10% glycerol, 150 mM NaCl, Roche Applied Science Complete protease inhibitor mixture, PhosStop, 50 mM *N*-ethylmaleimide). The lysate was centrifuged for 10 min at 17,000 × *g* at 4 °C. Equal protein amounts of each sample were then incubated with GFP-trap agarose beads from ChromoTek (Martinsried, Germany) that had been washed once in IP buffer (40 μl of 50% GFP-trap bead slurry per IP) and were incubated for 2 h at 4 °C. The beads were then washed three times with IP buffer by centrifugation at 2000 × *g* for 2 min at 4 °C. SILAC mixing was performed in the first wash. The beads were resuspended in 200 μl of PBS and transferred to a spin column (Pierce) and centrifuged dry at 500 *g* for 1 min. LDS sample buffer that had been preheated to 65 °C was then added and incubated at 65 °C for 5 min. The eluate was collected by centrifugation at 500 × *g* for 1 min at room temperature.

Immunofluorescence Microscopy—The cells were cultured on glass coverslips as described above. All of the subsequent steps are at 25 °C. The cells were then fixed with 3% paraformaldehyde in PBS. Fixed cells were washed with PBS, and free aldehyde groups were quenched with 50 mM glycine in PBS. The cells were then permeabilized using 1% Triton X-100 for 10 min followed by washing in PBS. Coverslips were processed for immunolabeling by blocking with 5% BSA in TBST. Primary antibodies were incubated on coverslips for 1 h in 5% BSA in TBST. The coverslips were washed by dipping 10 times into 500 ml of PBS. Primary antibodies were detected with Alexa Fluor 488- or Alex Fluor 594-conjugated secondary antibodies that were incubated on coverslips for 30 min in 5% BSA in TBST.

For paraffin-embedded paraformaldehyde-fixed mouse tissue sections (7 μm thick), deparaffinization was performed using Histo-Clear according to the manufacturer's instructions. Citrate-based unmasking solution was used for antigen retrieval according to manufacturer's instructions. The cells were permeabilized using 1% Triton X-100 for 15 min followed by washing in PBS. Four 15-min incubations with 1% sodium borohydride were used for autofluorescence reduction. The tissues were processed for immunolabeling by blocking with 5% donkey serum in PBS with 0.3% Triton X-100. Primary antibodies were incubated overnight in 5% donkey serum in PBS with 0.3% Triton X-100 at 4 °C. Primary antibodies were detected with Alexa Fluor 488- or Alex Fluor 594-conjugated secondary antibodies that were incubated for 1 h in 5% donkey serum in PBS with 0.3% Triton X-100 at 25 °C in the dark. Optical sections were analyzed by confocal microscopy on a Leica SP2 AOBS inverted microscope. Images were generated by the maximum projection of a z-stack taken from the top to the bottom of the cell monolayer. Contrast was adjusted for all images with the same settings.

RESULTS

Identification of Proteins Downstream of NEDD8-Cullin Pathway—Using a high specificity inhibitor of the NEDD8 E1 enzyme (MLN4924), we have interrogated the NEDD8 pathway in U2OS cells. Given that NEDD8 conjugation to Cullins is known to regulate their E3 ubiquitin ligase activity, we chose to identify those proteins whose expression increased after NEDD8 inhibition and therefore may be regulated downstream of the Cullins. We have used SILAC-based quantitative proteomics to perform a differential screen comparing the

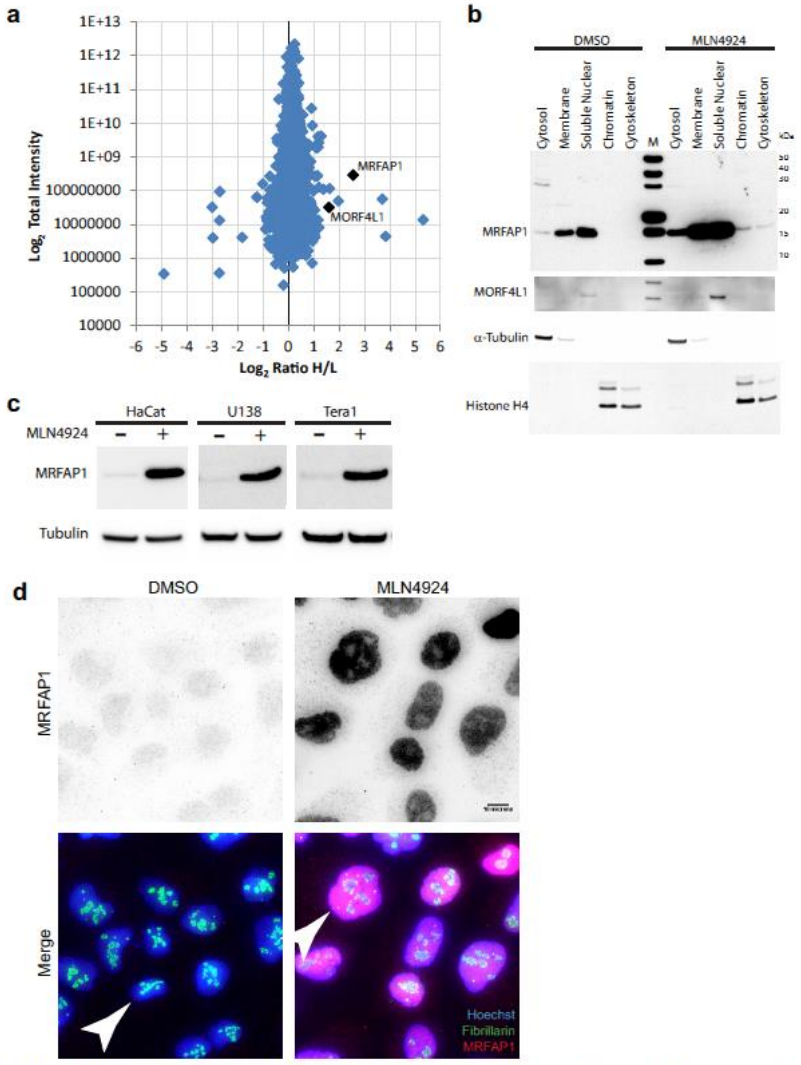


FIG. 1. **MRFAP1 and MORF4L1 are up-regulated in response to NEDD8 inhibition.** *a*, SILAC analysis of U2OS cells following NEDD8 inhibition. The abundance of ~2,500 proteins is indicated on the *y axis* using a $\log_2$ scale. The abundance of each protein indicated by the position of the dot on the *y axis* was determined by summing up all individual light and heavy peptide intensities detected for each protein. The relative fold decrease or increase is shown using the $\log_2$ ratio (heavy/light) on the *x axis*, with 1 μM MLN4924 treatment for 18 h (heavy) over DMSO only (light) ($n = 2$). *b*, immunoblotting of subcellular fractionation from U2OS cells treated with 1 μM MLN4924 treatment for 18 h or DMSO only ($n = 3$). *c*, immunoblotting of total cell lysates from the indicated cell types treated with 1 μM MLN4924 treatment for 18 h or DMSO only ($n = 2$). *d*, immunofluorescence microscopy of endogenous MRFAP1 in U2OS cells treated with 1 μM MLN4924 treatment for 18 h or DMSO only. The *white arrowheads* indicate change in nuclear size after MLN4924 treatment ($n = 3$). The *scale bar* indicates 10 μm.

relative abundance of proteins either in control, DMSO-treated U2OS cells or in U2OS cells treated for 18 h with 1 μM MLN4924 grown in medium containing either light (R0K0) or heavy (R10K8) stable isotope-labeled amino acids, respectively. Protein extracts from both the control and MLN4924-treated cells were mixed, fractionated, digested with trypsin into peptides, and analyzed by mass spectrometry on a LTQ Velos as described under "Experimental Procedures." Peptides were identified, and SILAC ratios were quantitated using MaxQuant, which showed that of >2,500 protein groups detected, 32 protein groups increase more than 2-fold in relative abundance after 18 h of MLN4924 treatment (Fig. 1*a* and

supplemental Table 1). Among these up-regulated proteins, MRFAP1 and MORF4L1 stood out from the top six most up-regulated proteins because they had been previously shown to be interaction partners (18) and to have fast turnover rates (4, 5). These two proteins were therefore selected for further analysis.

We confirmed the MS identification data using subcellular fractionation and protein blotting with antibodies specific for MRFAP1 and MORF4L1, which also showed that MRFAP1 and MORF4L1 are predominantly soluble nuclear proteins whose abundance increases following treatment of cells with MLN4924 (Fig. 1*b*). To determine whether the increase in the abundance of MRFAP1 in response to NEDD8 inhibition was conserved across different cell types, we also treated HaCaT (keratinocytes), U138MG (gliobastoma), and Tera-1 (pluripotent testicular teratoma) cells with MLN4924. Immunoblotting of total cell lysates from each of these cell types confirmed up-regulation of MRFAP1 levels in each case (Fig. 1*c*). To ensure that the antibody was specifically recognizing MRFAP1 and not other proteins in the extracts, we utilized siRNA knockdown of MRFAP1 in several cell lines, followed by both immunoblotting and indirect immunofluorescence microscopy (supplemental Fig. 1). In each case a major band at ~15 kDa, the predicted size for MRFAP1, was detected that was reduced after siRNA treatment with oligonucleotides specific for MRFAP1. These data showed that the antibody recognizes MRFAP1 specifically in multiple cell lines and would be useful for further analysis of the protein. To confirm the subcellular localization of MRFAP1, indirect immunofluorescence microscopy was performed on U2OS cells either with or without MLN4924 treatment. This showed MRFAP1 to have a nucleoplasmic localization that was increased in abundance after NEDD8 inhibition (Fig. 1*d*). A marked increase in nuclear size was also observed (Fig. 1*d*, *arrowheads*) and would be consistent with DNA rereplication, which is a known consequence of MLN4924 treatment (12).

MRFAP1 Is Stabilized by NEDD8 Inhibition and Co-expression of MORF4L1—To determine how MRFAP1 protein abundance was increased after NEDD8 inhibition, we tested whether the degradation of the protein was altered either with or without overnight MLN4924 treatment. A cycloheximide time course after overnight incubation of U2OS cells with either 1 μM MLN4924 or DMSO control treatment showed, as expected, an increase in MRFAP1 protein abundance upon NEDD8 inhibition, as well as a decrease in the rate of MRFAP1 degradation (Fig. 2*a*). Strikingly, analysis of DMSO-treated control cells showed MRFAP1 to have a very fast turnover rate, with almost no protein detected after 2 h of cycloheximide treatment. To confirm this result, we examined the turnover of LAP1-tagged, wild type MRFAP1 (supplemental Fig. 3*a*) expressed in an inducible stable U2OS cell line as described previously (22). Using an 8-h time course of cycloheximide treatment, we showed that the exogenous MRFAP1 had very similar degradation kinetics to endogenous

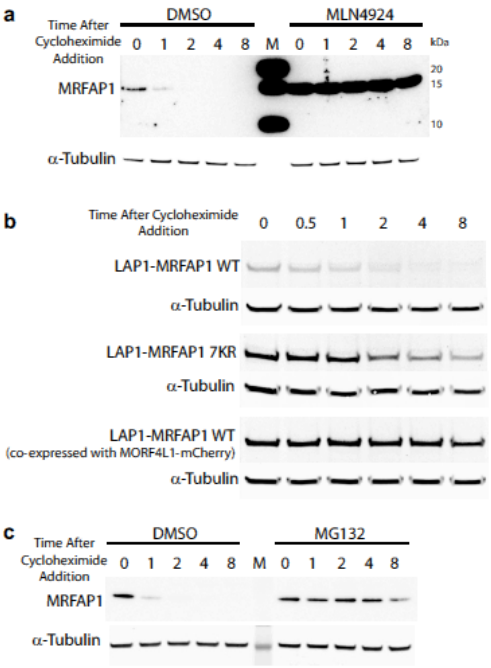


FIG. 2. **MRFAP1 is stabilized by NEDD8 inhibition.** *a*, immunoblotting of total cell lysates from U2OS cells treated with 1 μM MLN4924 for 18 h or DMSO only. A time course of cycloheximide treatment followed to examine protein degradation ($n = 3$). *b*, immunoblotting of total cell lysates from either LAP1-MRFAP1 wild type, LAP1-MRFAP1 7KR, or LAP1-MRFAP1-P2A-MORF4L1-mCherry after a time course of cycloheximide treatment to examine protein degradation ($n = 2$). *c*, immunoblotting of total cell lysates from U2OS cells treated with 10 μM MG132 or DMSO only. A time course of cycloheximide treatment was used to examine protein degradation. MG132 and cycloheximide were added at the same time for each time point ($n = 3$).

MRFAP1 (Fig. 2*b*). To determine whether this degradation was mediated via ubiquitin conjugation, we constructed a LAP1-tagged MRFAP1 mutant that was not competent for ubiquitin conjugation (7KR), which was expressed in a second inducible stable U2OS cell line just as for the wild type protein (Fig. 2*b*). In the 7KR mutant, all seven lysine residues in the MRFAP1 protein have been mutated to arginines (supplemental Fig. 3*b*), thereby preventing modification by ubiquitin, which must be linked to lysine residues. This fusion protein can, however, still be ubiquitinated within the LAP1 tag, which contains lysine residues (22). Regardless, the MRFAP1 7KR mutant was more stable than the wild type protein (Fig. 2*b*). To test whether it was the co-expression of MRFAP1 and MORF4L1 after MLN4924 treatment that helped stabilize each protein, we generated a third inducible stable cell line expressing each protein at close to equimolar levels. This cell line was generated using a new construct consisting of LAP1-tagged wild type MRFAP1 followed by a picornoviral ribo-

some skipping sequence (29) and then MORF4L1-mCherry (supplemental Fig. 3c). The exogenous LAP1-MRFAP1 in these cells showed similar increased stability in the presence of equimolar amounts of MORFL41 to that of endogenous MRFAP1 from cells treated with MLN4924 (Fig. 2*b*). To determine whether endogenous MRFAP1 degradation was mediated via the proteasome, we exposed U2OS cells to a cycloheximide time course with either 10 μM MG132 or DMSO control treatment. This showed, as expected, an increase in MRFAP1 protein abundance upon proteasome inhibition as well as a decrease in the rate of MRFAP1 degradation (Fig. 2*c*).

MRFAP1 Interacts with E3 Ubiquitin Ligases—Next we investigated which proteins may be controlling MRFAP1 and MORF4L1 turnover. To do this, we constructed inducible U2OS stable cell lines that express either wild type LAP1-MRFAP1 or wild type LAP1-MORF4L1 and used them to perform SILAC-based quantitation of the interaction partners for each of these proteins (22). In these triple SILAC experiments, we compared GFP-based immunoprecipitates from uninduced cells (light), doxycycline-induced (medium), and doxycycline-induced cells with simultaneous MLN4924 treatment (heavy). This experimental design allowed discrimination between strong interactors and contaminant proteins for each of the three conditions. Analysis of the LAP1-MRFAP1 immunoprecipitates revealed strong binding to previously identified interaction partners of MRFAP1, including MORF4L1 and MORF4L2, which were enriched in MLN4924-treated cells because of their increased abundance after NEDD8 inhibition (Fig. 3*a*). A number of ubiquitination-related proteins were also identified, as well as ubiquitin itself, including the HECT domain-containing E3 ligases UBR5 (EDD) and HUWE1, Cullin-4A/B, and one of its substrate-binding proteins VprBP, as well as the deubiquitination enzyme OTUB1. In addition, MRFAP1 showed interactions with coronin-1c, a WD40-repeat containing protein, which has previously been associated with F-actin binding but may have other roles with regard to MRFAP1 regulation. We were unable to detect any chromatin modification/remodeling enzymes in our MRFAP1 immunoprecipitates, which may indicate that MRFAP1, which binds to MORF4L1 via its MRG domain (20), blocks the MORF4L1 interaction with these complexes. Curiously, we consistently observed increased background binding of proteins from cells treated with MLN4924 as seen in Fig. 3*a* by the cloud of background binders offset from (0, 0). This phenomenon may relate to the effect of the drug on cellular architecture/cytoskeleton.

After comparing the SILAC ratios for each of these interactors, it was clear that VprBP and Cullin-4A/B only interacted strongly with MRFAP1 following MLN4924 treatment. This could be the result of an inactive cullin being unable to dissociate from its substrate because of an inability to carry out ubiquitination. To verify these data, we again immunoprecipitated LAP1-MRFAP1 from extracts of cells treated with MLN4924, separated the proteins by SDS-PAGE, and immunoblotted with antibodies specific for Cul4b, VprBP, and Cul4a. These data showed that Cul4b, but not Cul4a, was co-immunoprecipitated with MRFAP1 and confirmed our mass spectrometry quantitation, because the association of MRFAP1 with Cul4b was almost undetectable in untreated cells but was clearly seen from cells treated with either MLN4924 or MG132 (Fig. 3*b*). The interaction between

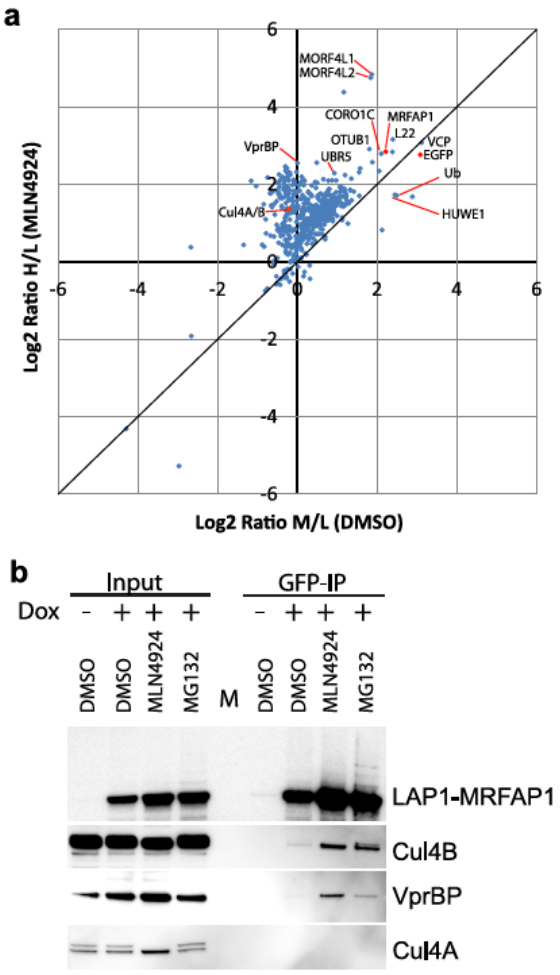


FIG. 3. **MRFAP1 interacts with ubiquitin E3 ligases.** *a*, SILAC GFP-IP analysis of U2OS cells stably expressing inducible LAP1-MRFAP1 wild type. GFP immunoprecipitates from uninduced (light), doxycycline-induced (medium), and doxycycline-induced cells with simultaneous MLN4924 treatment (heavy) were mixed and analyzed together by LC-MS/MS. The $\log_2$ ratio (medium/light) is indicated on the *x axis*. The $\log_2$ ratio (heavy/light) is indicated on the *y axis*. High ratios indicate high abundance of interaction compared with control. Proteins of interest are indicated ($n = 2$). *b*, immunoblotting of GFP-immunoprecipitated LAP1-MRFAP1. The cells were treated with 1 μM MLN4924 for 18 h or DMSO only ($n = 3$).

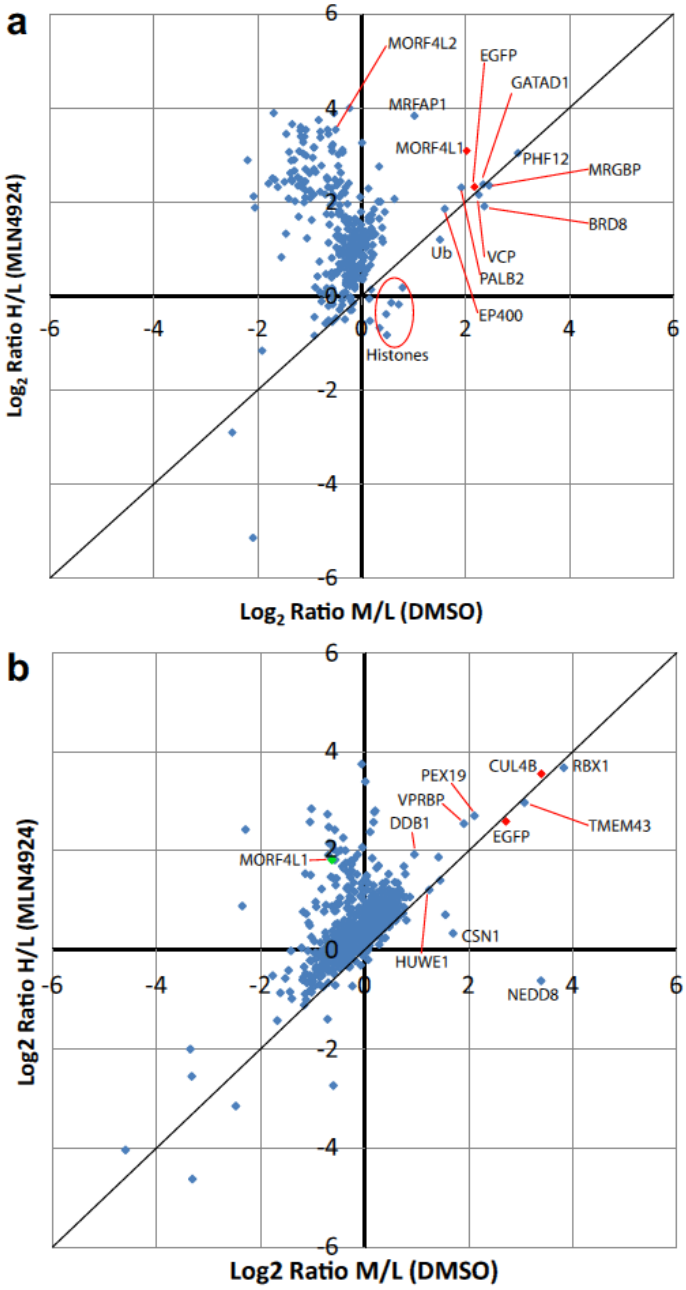


FIG. 4. **MRFAP1 is part of a complex protein-protein interaction network.** *a*, SILAC GFP-IP analysis of U2OS cells stably expressing inducible LAP1-MORF4L1. *b*, SILAC GFP-IP analysis of U2OS cells stably expressing inducible LAP1-CUL4B. GFP immunoprecipitates from uninduced (light), doxycycline-induced (medium), and doxycycline induced cells with simultaneous MLN4924 treatment (heavy) were mixed and analyzed together by LC-MS/MS. The $\log_2$ ratio (medium/light) is indicated on the *x axis*. The $\log_2$ ratio (heavy/light) is indicated on the *y axis*. High ratios indicate high abundance of interaction compared with control. Proteins of interest are indicated ($n = 1$).

MRFAP1 and VprBP was also confirmed and showed the strongest binding with MLN4924 treatment and weaker binding with MG132 treatment.

Analysis of the LAP1-MORF4L1 immunoprecipitates revealed strong binding to previously identified interaction partners of MORF4L1, including MRFAP1 and MRGBP, both of which bind to MORFL4 through its MRG domain (Fig. 4*a*). We observed strong binding to a number of other components of the NuA4 histone acetylation complex in addition to MRGBP, including BRD8 and EP400. Components of other chromatin-

modifying complexes were also seen, including GATAD1, a component of the JARID1A histone demethylase complex and PHF12, a component of the Sin3a histone deacetylase complex. Weaker interaction of MORFL41 with all of the core histones was also seen in control cells, but this was decreased after MLN4924 treatment. The recently identified MORF4L1-binding, DNA damage response protein PALB2 (FANCN) was also observed in our immunoprecipitates (30, 31). These data provide an overview of the variety of MORF4L1-containing complexes within U2OS cells and demonstrate our ability to observe several low abundance protein complexes.

To further clarify the role of Cul4B in the regulation of MRFAP1, we also carried out immunoprecipitation experiments to detect LAP1-CUL4B-interacting partner proteins in extracts from cells either with or without MLN4924 treatment (Fig. 4*b*). These data demonstrated how effective NEDD8 deconjugation was for Cul4b after MLN4924 treatment. Thus, we can clearly see NEDD8 association with Cul4b in untreated cells, as well as association with components of the deNEDDylating COP9 signalosome, such as CSN1 (10), as indicated by the high positive M/L ratio observed for these proteins. However, these interactions are largely abolished in MLN4924-treated cells as indicated by the negative H/L ratio observed for these proteins. We also observed interaction of Cul4b with the substrate-binding protein VprBP, which we also observed interacting with MRFAP1. However, MRFAP1 itself was not observed. MORF4L1 was observed interacting with Cul4b in MLN4924-treated cells, again indicating that Cul4b has an increased affinity for the MRFAP1-MORF4L1 complex after NEDD8 inhibition. TMEM43, which is a component of the inner nuclear membrane (32), showed a strong interaction that may be related to the nuclear localization of Cul4b in U2OS cells (Human Protein Atlas) (33). RBX1 and DDB1 are previously known components of Cul4b-containing protein complexes (34).

MRFAP1 Is Inversely Correlated with MRGBP Expression in Testis—Next we used a complementary microscopy-based strategy to investigate the expression pattern and associations of the interacting proteins identified above. Using the antibody specific for MRFAP1, which we previously validated by siRNA knockdown in a variety of cell types, we examined the tissue microarrays developed for the Human Protein Atlas project (33) to determine the expression pattern and subcellular localization of MRFAP1 in normal and cancerous human tissue (supplemental Tables 2 and 3). This analysis revealed that MRFAP1 had moderate nuclear expression in the spermatogonia of the testis, strong nuclear or cytoplasmic expression in ciliated epithelia, and strong cytoplasmic expression in neurons (Fig. 5*a*, nuclear staining indicated by *arrowheads*), whereas most other tissues had weak or negative staining (supplemental Table 2). The expression of MRFAP1 in spermatogonia may indicate that the protein plays some role in spermatogenesis and the initial stages of meiosis (supplemental Fig. 4).

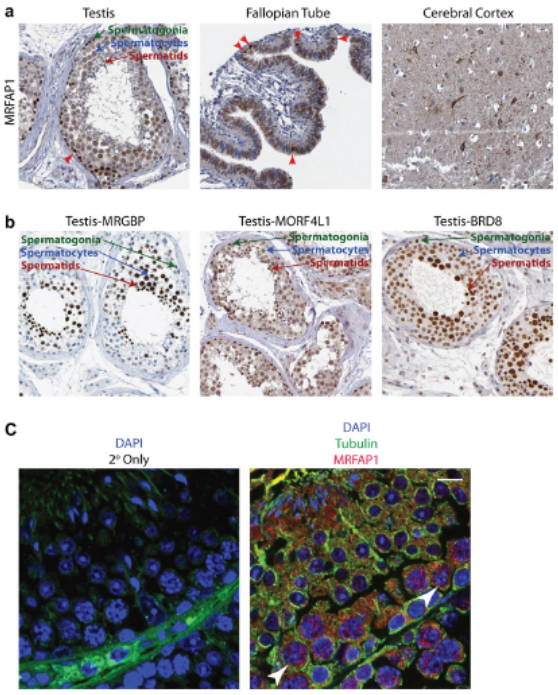


FIG. 5. **Human normal tissue expression pattern of MRFAP1.** *a*, immunohistochemistry imaging of endogenous MRFAP1 in selected tissues. The *brown color* indicates staining ($n = 3$). The *red arrows* indicate nuclear staining in spermatogonia of testis and ciliated epithelial cells of the fallopian tube. *b*, immunohistochemistry imaging of endogenous protein as indicated in testis tissue. The *brown color* indicates staining ($n = 3$). *c*, immunofluorescence microscopy of endogenous MRFAP1 in murine testis tissue ($n = 3$). The *red arrows* indicate nuclear staining in spermatogonia of testis. The *scale bar* indicates 10 μm.

To determine how the expression patterns of MORFL41, MRGBP, and BRD8 compared with that of MRFAP1, we examined the Human Protein Atlas testis tissue staining data (33), which utilized antibodies recognizing each of these proteins. This showed that MORF4L1 was detected at all cell stages of spermatogenesis, but that MRGBP and BRD8, two components of the NuA4 histone acetyltransferase complex, were highly enriched in spermatocytes and spermatids (Fig. 5*b*; Ref. 33; also see supplemental Fig. 4 for a description of spermatogenesis in the testis). Strikingly, when MRGBP expression in all tissues was examined, testis showed the strongest expression, and most other tissues were negative (33). In cancer tissues, MRFAP1 showed a very similar expression pattern to normal tissues, with the highest expression in glioblastoma and testicular cancer (supplemental Table 3). However, the MRGBP protein was largely undetectable in testicular cancer, in stark contrast to its expression in normal tissues. To confirm the subcellular localization of MRFAP1 in normal testis, we performed indirect immunoflu-

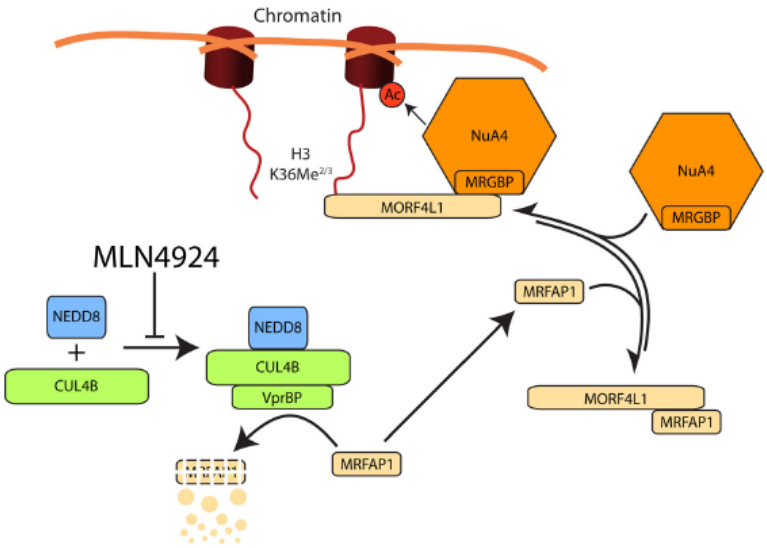


FIG. 6. **Model for the regulation of MRFAP1 by the NEDD8 pathway.** The NuA4 histone acetyltransferase complex is recruited to chromatin via the binding of MORF4L1 to tri/di-methylated lysine 36 on histone H3. Cells treated with MLN4924 have a defect in NEDD8 conjugation; therefore Cul4b would be less active in MRFAP1 degradation. An increase in free MRFAP1 protein displaces MRGBP from MORF4L1 and creates a stable MORF4L1-MRFAP1 complex.

orescence confocal microscopy on mouse testis sections (Fig. 5*c*). This analysis confirmed the nuclear localization of MRFAP1 in spermatogonia (Fig. 5*c*, *white arrowheads*). We conclude that in human seminiferous tubules of the testis, the MORF4L1 protein is expressed in all cells, but with a largely mutually exclusive co-expression of the alternate interacting proteins MRFAP1 and MRGBP.

DISCUSSION

In the present study, we have used SILAC-based quantitative mass spectrometry to characterize several nuclear proteins regulated by the NEDD8 pathway. We have identified MRFAP1 and MORF4L1 as being up-regulated in U2OS cells in response to inhibition of NEDD8 conjugation by treatment of cells with the inhibitor MLN4924. The up-regulation of MRFAP1 is mediated by an increase in the stability of the MRFAP1 protein after MLN4924 treatment. This stabilization could be enhanced because of the formation of a stable complex between MRFAP1 and MORF4L1. We further interrogated the protein-protein interaction network formed by these proteins and identified the E3 ubiquitin ligase Cul4b complex as binding to MRFAP1. This provides a putative mechanism to link the NEDD8-regulated Cul4b protein and stabilization of MRFAP1 (Fig. 6). We also saw that MRFAP1 and MRGBP could not be co-immunoprecipitated and therefore may bind in a mutually exclusive manner to MORFL41. We infer that MRFAP1 could therefore act as a regulator of MORF4L1 activity by changing its binding partners.

The proteomic analysis we present here highlighted only a subset of proteins whose expression was up-regulated by 18 h of MLN4924 treatment. The short period of NEDD8 conjugation inhibition used in this study may favor the detection of proteins that are rapidly degraded, downstream of the NEDD8-cullin E3 ubiquitin ligase pathways. A recent study that examined in detail the transcriptional and proteome response to MLN4924 treatment in human A375 melanoma cells reported a larger number of proteins whose abundance was altered (14). This may reflect differences in the cell types studied and/or experimental conditions used. Regardless, Liao *et al.* (14) also observe the increased abundance of MRFAP1 and MORF4L1 in human A375 melanoma cells after MLN4924 treatment, which adds further evidence to the universal response of cells to up-regulate these proteins after NEDD8 conjugation inhibition.

Analysis of the human tissue expression pattern of MRFAP1 showed it to be expressed in only a few cell types, including spermatogonia, ciliated epithelial cells, and neurons. Closer examination of the testis expression pattern for MRFAP1 and MRGBP indicated an anti-correlated pattern, suggesting that recruitment of NuA4 histone acetyltransferase activity to chromatin via MORFL41 might be important for later stages of spermatogenesis. This is because only the cells in later stages of spermatogenesis showed co-expression of MRGBP and MORF4L1 but not MRFAP1. The MORF4L1-MRGBP modification complex could be inhibited in spermatogonia because of high MRFAP1 expression, which we suggest competes with MRGBP for binding to MORF4L1. We propose that this may block the hyperacetylation pathway in spermatogonia to maintain a germ stem cell population.

Previous studies have shown in a number of animals, including rats and mice, that hyperacetylation of histone H4 is critical for spermatogenesis (35, 36). These studies showed that hyperacetylation of histone H4 leads to chromatin decondensation, which is thought to allow for histone replacement by protamines and thereby facilitate the extensive chromatin compaction needed in the small sperm nu-

cleus (supplemental Fig. 4). We propose that the NuA4 histone acetyltransferase complex, which contains MRGBP and BRD8, plays a critical role in spermatogenesis by mediating the hyperacetylation of histone H4. One of the other MORF4L1-interacting proteins we observed was EP400, which as part of the human NuA4 complex may also play a role in the exchange of histone H2A for histone H2A.Z during spermatogenesis (37).

The MRFAP1 protein is highly conserved and only present in mammals, consistent with it playing a specific role in the regulation of MORF4L1 protein complexes (supplemental Fig. 2). MRFAP1 may thus be expressed in spermatogonia to maintain normal levels of histone modification by negatively regulating recruitment of the NuA4 complex to chromatin (Fig. 6).

Analysis of the MRFAP1 domain structure shows that it is a relatively simple, modular protein, with an N-terminal MRG-binding domain followed by a flexible linker region containing a number of glycine residues and a C-terminal coiled coil region that could mediate homodimerization of the MRFAP1 protein (38–40). The dimerization capability of MRFAP1 could reflect the fact that the MRG domain of MORF4L1 is a dimer, the structure of which has been solved by x-ray crystallography (20). We propose that a homodimer of MRFAP1 may bind the homodimer of MORF4L1, forming a high affinity interaction between the two dimers. The abundance of the MORF4L1-MRFAP1 complex, compared with the MORF4L1-MRGBP complex, would then be mediated, at least in part, by the probability of an encounter between each of these proteins and therefore is related to the abundance and localization of each protein (19).

Our studies here using quantitative proteomics analyses have revealed new evidence for the functional role of NEDD8 modification in controlling gene expression during cell differentiation in mammals. This has opened up new avenues to investigate mechanisms regulating histone acetylation and its link with spermatogenesis. In view of the tissue specificity of MRFAP1 and MRGBP expression, it is possible that these proteins regulate the hyperacetylation of histone H4 during spermatogenesis via a pathway dependent upon binding stoichiometry. In the context of cancer, this pathway may be involved by subtly altering the histone modifications of chromatin, thereby altering gene expression to favor cancer cell proliferation. Our data provide evidence for a combinatorial mechanism affecting cell-specific regulation of chromatin modification that involves changing the balance between MRFAP1-MORF4L1-MRGBP interactions. In future, it will be of interest to characterize further the impact of this pathway in disease states, such as cancer, that are influenced by alterations in histone acetylation and related modifications.

* This work was supported by Wellcome Trust Grants 083524/Z/07/Z, 073980/Z/03/Z, 08136/Z/03/Z, and 0909444/Z/09/Z; EU FP7 Prospects Network Grant HEALTH-F4–2008-201648; and European Union FP7 Epigenesis Network Grant HEALTH-F4-2010-257082. The Human Protein Atlas project was funded by the Knut and Alice Wallenberg Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] This article contains supplemental Tables 1–3 and Figs. 1–4.

§ Association for International Cancer Research Fellow. Present address: Universités Montpellier 2 et 1, Centre de Recherche de Biochimie Macromoléculaire, CNRS UMR 5237, Montpellier, France.

‖ Wellcome Trust Principal Research Fellow. To whom correspondence should be addressed: Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow St., Dundee, UK. Tel.: 44-01382385473; Fax: 44-01382385695.

REFERENCES

1. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1,** 376–386
2. Cox, J., and Mann, M. (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80,** 273–299
3. Lam, Y. W., Lamond, A. I., Mann, M., and Andersen, J. S. (2007) Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr. Biol.* **17,** 749–760
4. Boisvert, F. M., Ahmad, Y., Gierlinski, M., Charriere, F., Lamond, D., Scott, M., Barton, G., and Lamond, A. I. (2011) A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol. Cell. Proteomics*, 10.1074/mcp.M111.011429
5. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* **473,** 337–342
6. Hori, T., Osaka, F., Chiba, T., Miyamoto, C., Okabayashi, K., Shimbara, N., Kato, S., and Tanaka, K. (1999) Covalent modification of all members of human cullin family proteins by NEDD8. *Oncogene* **18,** 6829–6834
7. Pan, Z. Q., Kentsis, A., Dias, D. C., Yamoah, K., and Wu, K. (2004) Nedd8 on cullin: Building an expressway to protein destruction. *Oncogene* **23,** 1985–1997
8. Osaka, F., Kawasaki, H., Aida, N., Saeki, M., Chiba, T., Kawashima, S., Tanaka, K., and Kato, S. (1998) A new NEDD8-ligating system for cullin-4A. *Gene Dev.* **12,** 2263–2268
9. Huang, D. T., Ayrault, O., Hunt, H. W., Taherbhoy, A. M., Duda, D. M., Scott, D. C., Borg, L. A., Neale, G., Murray, P. J., Roussel, M. F., and Schulman, B. A. (2009) E2-RING expansion of the NEDD8 cascade confers specificity to cullin modification. *Mol. Cell* **33,** 483–495
10. Lyapina, S., Cope, G., Shevchenko, A., Serino, G., Tsuge, T., Zhou, C., Wolf, D. A., Wei, N., Shevchenko, A., and Deshaies, R. J. (2001) Promotion of NEDD8-CUL1 conjugate cleavage by COP9 signalosome. *Science* **292,** 1382–1385
11. Read, M. A., Brownell, J. E., Gladysheva, T. B., Hottelet, M., Parent, L. A., Coggins, M. B., Pierce, J. W., Podust, V. N., Luo, R. S., Chau, V., and Palombella, V. J. (2000) Nedd8 modification of Cul-1 activates SCF beta(TrCp)-dependent ubiquitination of I kappa B alpha. *Mol. Cell. Biol.* **20,** 2326–2333
12. Soucy, T. A., Smith, P. G., Milhollen, M. A., Berger, A. J., Gavin, J. M., Adhikari, S., Brownell, J. E., Burke, K. E., Cardin, D. P., Critchley, S., Cullis, C. A., Doucette, A., Garnsey, J. J., Gaulin, J. L., Gershman, R. E., Lublinsky, A. R., McDonald, A., Mizutani, H., Narayanan, U., Olhava, E. J., Peluso, S., Rezaei, M., Sintchak, M. D., Talreja, T., Thomas, M. P., Traore, T., Vyskocil, S., Weatherhead, G. S., Yu, J., Zhang, J., Dick, L. R., Claiborne, C. F., Rolfe, M., Bolen, J. B., and Langston, S. P. (2009) An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **458,** 732–736
13. Jia, L., Li, H., and Sun, Y. (2011) Induction of p21-dependent senescence by an NAE inhibitor, MLN4924, as a mechanism of growth suppression. *Neoplasia* **13,** 561–569
14. Liao, H., Liu, X. J., Blank, J. L., Bouck, D. C., Bernard, H., Garcia, K., and Lightcap, E. S. (2011) Quantitative proteomic analysis of cellular protein modulation upon inhibition of the NEDD8-activating enzyme by MLN4924. *Mol. Cell. Proteomics*, 10.1074/mcp.M111.009183

15. Doyon, Y., and Côté, J. (2004) The highly conserved and multifunctional NuA4 HAT complex. *Curr. Opin. Genet. Dev.* **14,** 147–154
16. Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006) Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res.* **34,** 6621–6628
17. Pardo, P. S., Leung, J. K., Lucchesi, J. C., and Pereira-Smith, O. M. (2002) MRG15, a novel chromodomain protein, is present in two distinct multiprotein complexes involved in transcriptional activation. *J. Biol. Chem.* **277,** 50860–50866
18. Leung, J. K., Berube, N., Venable, S., Ahmed, S., Timchenko, N., and Pereira-Smith, O. M. (2001) MRG15 activates the B-myb promoter through formation of a nuclear complex with the retinoblastoma protein and the novel protein PAM14. *J. Biol. Chem.* **276,** 39171–39178
19. Beckett, D. (2004) Functional switches in transcription regulation: Molecular mimicry and plasticity in protein-protein interactions. *Biochemistry* **43,** 7983–7991
20. Zhang, P., Zhao, J., Wang, B., Du, J., Lu, Y., Chen, J., and Ding, J. (2006) The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. *Protein Sci* **15,** 2423–2434
21. Tominaga, K., Magee, D. M., Matzuk, M. M., and Pereira-Smith, O. M. (2004) PAM14, a novel MRG- and Rb-associated protein, is not required for development and T-cell function in mice. *Mol. Cell. Biol.* **24,** 8366–8373
22. Torres, J. Z., Miller, J. J., and Jackson, P. K. (2009) High-throughput generation of tagged stable cell lines for proteomic analysis. *Proteomics* **9,** 2888–2891
23. Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., Georgieva, M., Azzag, K., Robert, M. C., Ahmad, Y., Neel, H., Lamond, A. I., and Bertrand, E. (2010) HSP90 and its R2TP/prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol. Cell* **39,** 912–924
24. Boisvert, F. M., Lam, Y. W., Lamont, D., and Lamond, A. I. (2010) A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol. Cell. Proteomics* **9,** 457–470
25. Wessel, D., and Flügge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138,** 141–143
26. Bereman, M. S., Egertson, J. D., and MacCoss, M. J. (2011) Comparison between procedures using SDS for shotgun proteomic analyses of complex samples. *Proteomics* **11,** 2931–2935
27. Haas, W., Faherty, B. K., Gerber, S. A., Elias, J. E., Beausoleil, S. A., Bakalarski, C. E., Li, X., Villén, J., and Gygi, S. P. (2006) Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol. Cell. Proteomics* **5,** 1326–1337
28. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372
29. Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F., and Vignali, D. A. (2004) Correction of multi-gene deficiency in vivo using a single "self-cleaving " 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22,** 589–594
30. Sy, S. M., Huen, M. S., and Chen, J. (2009) MRG15 is a novel PALB2-interacting factor involved in homologous recombination. *J. Biol. Chem.* **284,** 21127–21131
31. Hayakawa, T., Zhang, F., Hayakawa, N., Ohtani, Y., Shinmyozu, K., Nakayama, J., and Andreassen, P. R. (2010) MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. *J. Cell Sci.* **123,** 1124–1130
32. Bengtsson, L., and Otto, H. (2008) LUMA interacts with emerin and influences its distribution at the inner nuclear membrane. *J. Cell Sci.* **121,** 536–548
33. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Björling, L., and Ponten, F. (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol* **28,** 1248–1250
34. Jackson, S., and Xiong, Y. (2009) CRL4s: The CUL4-RING E3 ubiquitin ligases. *Trends Biochem. Sci* **34,** 562–570
35. Christensen, M. E., and Dixon, G. H. (1982) Hyperacetylation of histone-H4 correlates with the terminal, transcriptionally inactive stages of spermatogenesis in rainbow trout. *Dev. Biol.* **93,** 404–415
36. Govin, J., Caron, C., Lestrat, C., Rousseaux, S., and Khochbin, S. (2004) The role of histones in chromatin remodelling during mammalian spermiogenesis. *Eur. J. Biochem.* **271,** 3459–3469
37. Auger, A., Galarneau, L., Altaf, M., Nourani, A., Doyon, Y., Utley, R. T., Cronier, D., Allard, S., and Côté, J. (2008) Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. *Mol. Cell. Biol.* **28,** 2257–2270
38. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc* **4,** 363–371
39. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Clustal-W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680
40. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview version 2: A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191

In order to cite this article properly, please include all of the following information: Larance, M., Kirkwood, K. J., Xirodimas, D. P., Lundberg, E., Uhlen, M., and Lamond, A. I. (2012) Characterization of MRFAP1 Turnover and Interactions Downstream of the NEDD8 Pathway. *Mol. Cell. Proteomics* 11(3):M111.014407. DOI: 10.1074/mcp.M111.014407.